INTERACTION OF PROTEINS WITH SMALL MOLECULES AND PEPTIDES

Uttam Pal

Indian Institute of Chemical Biology

2016

Interaction of Proteins with Small Molecules and Peptides

A thesis submitted for the degree of Doctor of Philosophy of the Jadavpur University on 22 February 2016

by
Uttam Pal, INSPIRE Fellow
Structural Biology & Bioinformatics Division
CSIR-Indian Institute of Chemical Biology, Kolkata

Supervisor
Dr. Nakul Chandra Maiti, Senior Scientist
Structural Biology & Bioinformatics Division
CSIR-Indian Institute of Chemical Biology, Kolkata

CERTIFICATE FROM THE SUPERVISOR

This is to certify that the thesis entitled "Interaction of Proteins with Small Molecules and Peptides" submitted by Uttam Pal, who got his name registered on April 29, 2014 (Reference No. D-7/Sc/358/14, Index No. 85/14/Chem./23) for the award of Ph.D. (Science) degree of Jadavpur University, is absolutely based upon his own work under the supervision of Dr. Nakul Chandra Maiti and that neither this thesis nor any part of it has been submitted for either any degree/diploma or any other academic award anywhere before.

Nakul Chandra Maiti 22-02-2016

.....

(Signature of the Supervisors and date with official seal)

Dr. Nakul Chandra Maiti
वरिष्ट वैज्ञानिक / Senior Scientist
संरचनागत जीवविज्ञान और जैवसुचना प्रभाग
Structural Biology & Bioinformatics Division
भारतीय रासायनिक जीवविज्ञान संस्थान
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
Indian Institute of Chemical Biology
(Council of Scientific & Industrial Research)
कोलकाता / Kolkata-700032

DECLARATION BY THE AUTHOR

This thesis is submitted to the Jadavpur University in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. No data provided and/or analysis carried out by collaborators has been included in this work.

Parts of this thesis have been published by the author:

i. Uttam Pal, and Nakul Chandra Maiti,
Allostery and Druggability Prediction by Molecular Docking,
Journal of Proteins and Proteomics **6**, 133 (2015).

ii. Uttam Pal, Mritunjoy Maity, Nitin Khot, Swagata Das, Supriya Das, Sandip Dolui, and Nakul Chandra Maiti
Statistical Insight into the Binding Regions in Disordered Human Proteome,
Journal of Proteins and Proteomics (In press).

iii. Uttam Pal, Sumit Kumar Pramanik, Baisali Bhattacharya, Biswadip Banerji, Nakul Chandra Maiti,
Binding Interaction of a Novel Fluorophore with Serum Albumins: Steady State Fluorescence Perturbation and Molecular Modeling Analysis,
SpringerPlus **4**, 548 (2015).

iv. Uttam Pal, Sudeshna Sen, and Nakul Chandra Maiti,
C_α–H Carries Information of a Hydrogen Bond Involving the Geminal Hydroxyl Group: A Case Study with a Hydrogen-Bonded Complex of 1,1,1,3,3,3-Hexafluoro-2-propanol and Tertiary Amines,
The Journal of Physical Chemistry A **118**, 1024–1030 (2014).

Date: 22 February 2016
Place: Kolkata

Uttam Pal
.....
(Signature of the Author)

To K.B.

SYNOPSIS

Understanding the molecular interaction of proteins with small molecules and peptides is essential in drug discovery. This thesis explores the importance of protein small molecule interactions in biology and its implications in rational drug development. Using combined theoretical and experimental chemistry, author tried to device new approaches or build upon the existing tools to study the interactions of proteins with small molecules and peptides. New algorithm has been developed based on the existing molecular docking methods to improve the specificity in docking prediction and reduce the false positive results (hits) in virtual high throughput screening. In recent years allosteric drugs have gained much attention because of the specificity it can achieve due to the evolutionarily diverse binding sites on the target proteins. The author developed a pattern based approach using molecular docking tools to detect the allosteric binding sites on structured protein targets and analyzed the druggability of these allosteric binding sites. Detections of binding sites in the intrinsically disordered proteins was also explored using the sequence based analysis tools. Statistical distribution of binding sites in a library of disordered proteins was characterized. A fluorescent probe molecule has been developed to study the microenvironment of protein binding sites, which showed increased quantum yield, anisotropy and higher energy emission upon binding to a hydrophobic site of the target protein. Author studied even deeper into the finer details of molecular interactions such as the hydrogen bonding and its strength on model system (alcohol amine complexes) using isotope labeling, solution state nuclear magnetic resonance spectroscopy and quantum mechanics with implications in protein small molecule interactions.

CONTENTS

BACKGROUND

"For we know in part..."
1 Corinthians 13:9

Importance of protein-small molecule interaction. A number of different types of molecular interactions enable life, which include the interactions between proteins, proteins and nucleic acids, and proteins and small molecules (McFedries et al., 2013). In all these three types of interactions the common factor is protein. Proteins are the most common molecules found in cells. They come in various forms and shapes and carry out diverse functions specified by the information encoded in genes (Au et al., 2008), including the regulation of genetic information. Elucidating the interactions of proteins with other proteins/nucleic acids/small molecules and understanding how they control biology is a major scientific goal (McFedries et al., 2013).

Enzymes and their modulators. Best known role of proteins are as enzymes, which catalyze chemical reactions. Enzymes are highly specific and accelerates only one or a few chemical reactions. There are more than 5,000 biochemical reaction types are known to be catalyzed by enzymes and the substrate ranges from small molecules to biopolymers (Schomburg et al., 2013). The highly sophisticated metabolic network of enzyme catalyzed reactions are often regulated by enzyme inhibition. End product of a biochemical reaction may inhibit the key rate limiting enzyme of the pathway in a feedback mechanism, thus, causing the production of the substance to slow down or stop when there is sufficient amount. Major biochemical pathways such as the citric acid cycle make use of this mechanism. Since inhibitors can modulate enzyme function, they are often used as drugs. Many such drugs are reversible competitive inhibitors that resemble the enzyme's native substrate or the transition state structure of the substrate (Schramm, 2005, 2007, 2013). A well know example is methotrexate, which is used in cancer chemotherapy (Jolivet et al., 1983). Methotrexate binds to the dihydrofolate reductase enzyme and prevents the binding of its original substrate, folic acid, to the enzyme. Other

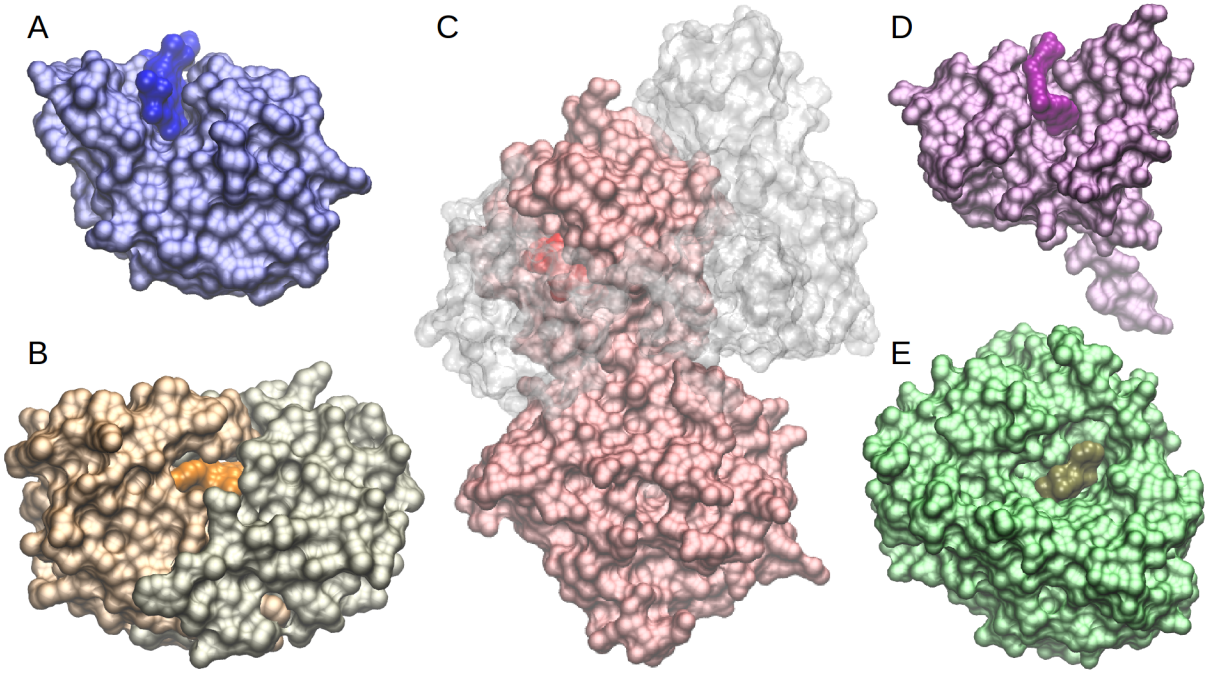


Figure 1: Enzymes bound with small molecule ligands. (A) Lysozyme (PDB: 1LZR) bound with hexa-N-acetyl-chitohexaose oligosaccharide. Lysozyme was the second protein structure and the first enzyme structure to be solved by X-ray diffraction methods. With more than a thousand structures in protein data bank (PDB), lysozyme is one of the most studied enzyme till date. (B) HIV-1 protease (PDB: 4DQB) complexed with darunavir. With its integral role in HIV replication, HIV protease has been a prime target for antiretroviral drug discovery. (C) Gelatinase B (matrix metalloproteinase 9) complexed with a hydroxamate based inhibitor. It plays a central role in tumor progression, from angiogenesis, to stromal remodeling, and ultimately metastasis. PDB files 4WZV, 1ITV and 1CK7 (gelatinase A) were used to create this illustration. (D) Dihydrofolate reductase (DHFR) inhibited by chemotherapeutic methotrexate which prevents binding of its substrate, folic acid. (PDB: 4QI9). DHFR is an attractive pharmaceutical target due to its pivotal role in DNA precursor synthesis. (E) The structure of a complex between penicillin G and the Streptomyces transpeptidase generated from PDB 1PWC. Penicillin binding proteins, which are essential for bacterial cell wall biogenesis are attractive target for drug development.

well-known examples include protease inhibitors used to treat retroviral infections such as HIV (human immunodeficiency virus) (Deeks SG et al., 1997). Most common examples of inhibitors that is used as drugs are penicillin and aspirin. Penicillin and its more potent derivatives such as amoxicillin, ampicillin etc. are used to treat a number of bacterial infections (Lee NS et al., 2001). They act as an irreversible inhibitor of the enzyme transpeptidase, which is needed by bacteria to make their cell walls. Aspirin, on the other hand, inhibits the cyclooxygenase enzymes that

produce the inflammation messenger prostaglandin, thus, provides relief from pain and inflammation (Sneader, 2000). Small molecule interaction with enzymes may be poisonous as well. For example, cyanide binds irreversibly with the copper and iron in the active site of the enzyme cytochrome c oxidase and blocks cellular respiration. Figure 1 illustrates some important enzymes including gelatinase B. Our work on the drug discovery for gelatinase B is discussed in the chapter 1 of this thesis (Rudra et al., 2012). The other enzyme for which we have developed inhibitors is the macrophage migration inhibitory factor (MIF), which is an important regulator of innate immunity (Alam et al., 2011, 2012).

Receptors and their ligands. Proteins are involved in the process of cell signaling and signal transduction. Membrane proteins that act as receptors bind to a signaling molecule and induce a biochemical response in the cell. Many such receptors have extracellular ligand binding domain exposed on the outer surface of the cell. When ligand molecule binds on it, a conformational change is triggered, which may lead to the opening of a transmembrane channel or activate the intracellular effector domain. For example, the neurotransmitter GABA (γ-aminobutyricacid) can activate a cell surface receptor that is part of an ion channel. GABA binding to the extracellular domain of a $GABA_A$ receptor on a neuron opens a chloride-selective ion channel that is part of the receptor. Opening of this channel allows negatively charged chloride ions to move into the neuron, which inhibits the ability of the neuron to produce action potentials (Kuffler and Edwards, 1958; Miller and Aricescu, 2014). However, for many cell surface receptors, ligand-receptor interactions are not directly linked to the cell's response, rather a cascade of reactions with other proteins inside the cell is initiated before the ultimate physiological effect of the ligand on the cell's behavior is produced. Often, it includes small molecule mediators inside the cells called the second messenger such as IP_3 (Inositol trisphosphate) or cAMP (cyclic adenosine monophosphate). G protein coupled receptors (GPCR) are well known examples for such signal transduction. The ligands that bind and activate these receptors include odors, pheromones, hormones, and neurotransmitters, and vary in size from small molecules to peptides to large proteins. G protein coupled receptors are involved in

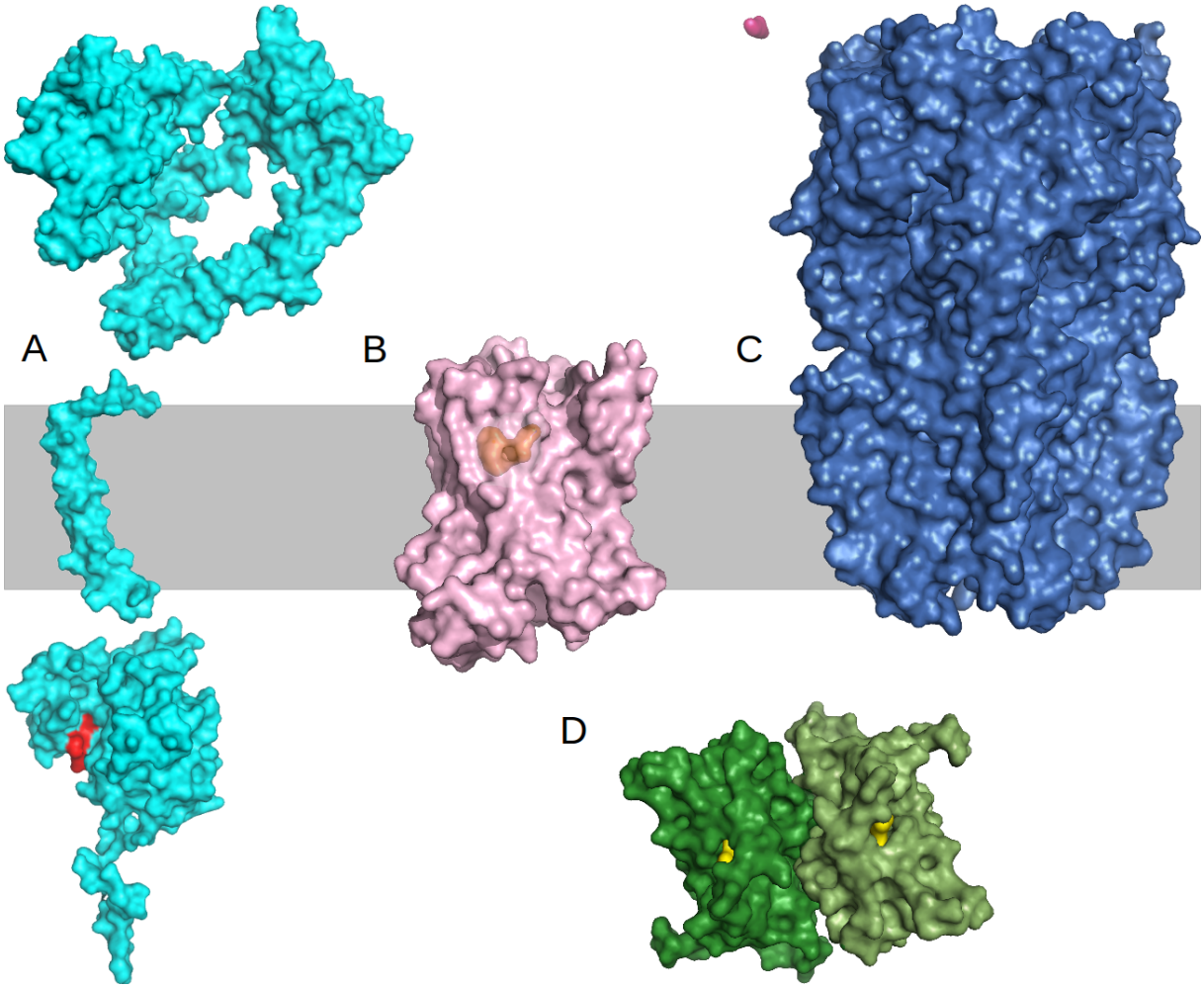


Figure 2: Receptor proteins with small molecule ligands. (A) Epidermal growth factor receptor (EGFR) bound to the inhibitor erlotinib. Mutation in EGFR have been associated with a number of cancers, including lung cancer and breast cancers. Several PDB files were needed to create this illustration, including 1nql, 2jwa, 1m17. The cell membrane is shown schematically in gray. (B) β2-adrenergic receptor bound to the inhibitor carazolol (PDB: 2RH1). It is a part of a large class of similar proteins, collectively known as G-protein-coupled receptors (GPCRs). Many widely-used drugs, such as Prozac, Claritin, and Zoloft, act by binding to proteins involved in GPCR signaling. (C) GABAA receptor (PDB: 4COF) and the signaling molecule γ-aminobutyricacid (shown in pink). It is a target of the benzodiazepine class of tranquilizer drugs. (D) Ligand binding domain of estrogen receptor with breast cancer medication tamoxifen (shown in yellow). PDB: 3ERT.

many diseases, and are also the target of approximately 30% of all modern medicinal drugs (Overington et al., 2006). Small molecule interference with signal transduction has historical importance as well. It was exploited during the World War II to produce the weapons of mass destruction. Tabun, Sarin and Soman are the three infamous organophosphates that disrupt the mechanism by which nerves

transfer messages to organs by blocking acetylcholinesterase, an enzyme that normally destroys the neurotransmitter acetylcholine (Millard et al., 1999; Sanson et al., 2009). Figure 2 shows some of the important receptor proteins including epidermal growth factor receptor (EGFR) for which we have developed an inhibitor (Bhowmik et al., 2013). Overexpression of EGFR is the most commonly observed cancer associated misregulation in EGFR signaling and correlates in a number of cancers including breast, ovarian, and head and neck (Roberts et al., 2002). The computational aspects of the EGFR inhibitor development has been highlighted in the chapter 1 of this thesis.

Transporters and their substrates. There are transport proteins that serve the function of moving other materials within an organism. Transport proteins are vital to the growth and life of all living things. There are several kinds of transport proteins including the secreted and membrane proteins, a few of which are illustrated in the figure 3. Transmembrane proteins that alter the permeability of the cell membrane to small molecules and ions serve as ligand transport proteins. One such example is gastric hydrogen potassium pump, found in the parietal cells of gastric mucosa. It pumps the protons into the gastric lumen from inside the parietal cells in exchange of potassium ions, thus, acidifying the stomach. It is also responsible for chronic inflammation to peptic ulcer. Omeprazole is a small molecule that can inhibit this proton pump and used to treat a wide range of gastrointestinal tract ailments including gastritis and peptic ulcer disease (Olbe et al., 2003). Besides the membrane transporters, many ligand transport proteins bind particular small biomolecules and transport them to other locations in the body of a multicellular organism. These proteins show high binding affinity when the ligand is present in high concentrations, but release the ligand when it is present at low concentrations in the target tissues. The canonical example is the hemoglobin, which transports oxygen from lungs to all parts of the body. High affinity binding of ultrafine carbon nanoparticles with hemoglobin and its potential health hazard has been addressed in one of our recent publications (Banerji et al., 2014). Another most abundant transport protein of blood plasma is serum albumin. It acts as a plasma carrier by non-specifically binding several hydrophobic steroid hormones

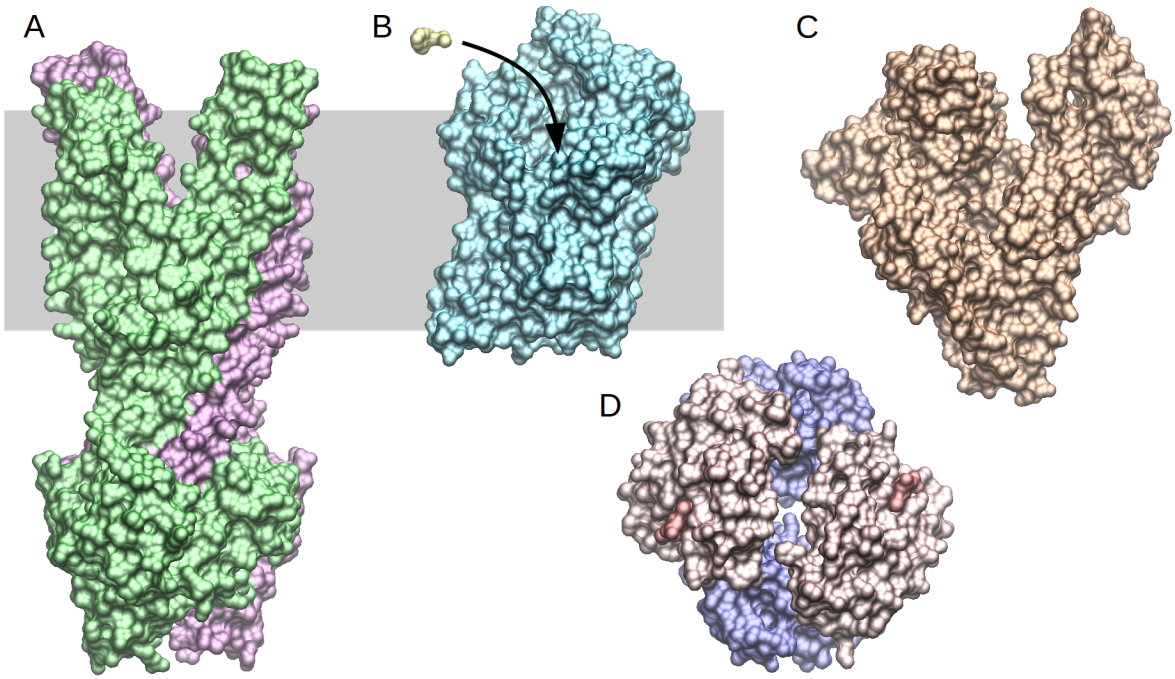


Figure 3: Transporter proteins. (A) Bacterial multidrug resistant transporter Sav1866 from PDB entry 2onj, is found in Staphylococcus bacteria. These transporters find drugs that try to gain entry through a cell membrane and transport them back outside. (B) Dopamine neurotransmitter transporter with an antidepressant drug nortriptyline (PDB: 4M48). Drugs that block the action of the transporter cause the neurotransmitter to remain in the synaptic cleft longer than normal. Antidepressant drugs (such as the one shown here in yellow) take advantage of this by blocking the transporters. Drugs of abuse like cocaine also block the action of these transporters. (C) Serum albumin, shown here from PDB entry 1E78 , is the carrier of fatty acids in the blood. Serum albumin also binds to many other water-insoluble molecules, such as drug molecules, and can strongly affect the way they are delivered through the body. (D) Hemoglobin (PDB: 1A3N) is a transporter of oxygen. It also binds to other molecules. In particular, hemoglobin binds to some metabolites, such as 2,3-bisphosphoglycerate which affects its oxygen binding affinity.

and as a transport protein for hemin and fatty acids. Drugs that enters circulation also binds to serum albumins, which may improve the bioavailability and retention time of the drug. The serum protein binding is generally nonspecific and reversible and it can influence the drug's biological half-life in the body. The bound portion may act as a reservoir from which the drug is slowly released as the unbound form. Since the unbound form is being metabolized and/or excreted from the body, the bound fraction is gradually released in order to maintain equilibrium. Serum albumin is one of the most extensively studied proteins by fluorescence spectroscopy. Some of our previous works on protein small molecule interactions

were performed on serum albumin and its bovine orthologoue (Banerjee et al., 2012; Banerji et al., 2013a; Maity et al., 2014; Ray et al., 2012). Our study of the micro-environment of the binding sites in protein, as highlighted in the chapter 4 of this thesis, has also been carried out on the serum proteins (Pal et al., 2015).

Building blocks and their demolishers. However, the most abundant class of protein in nature is the structural proteins, which confer stiffness and rigidity to otherwise-fluid biological components (Kielty et al., 2002). Structural proteins are generally fibrous; for example collagen and elastin are critical components of connective tissue such as extracellular matrix and cartilage (Halper and Kjaer, 2014). Some soluble globular proteins, for example, actin and tubulin also play structural functions upon polymerization. Actin polymerize to form long, stiff fibers

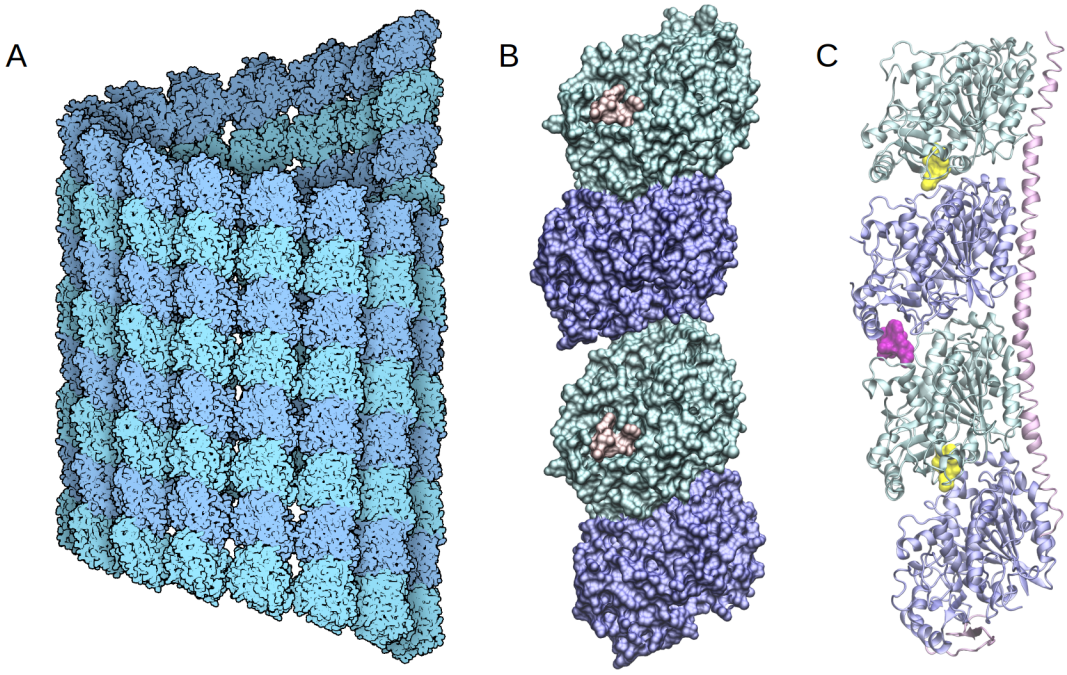


Figure 4: Structural protein tubulin with inhibitors of polymerization. (A) Alpha and beta tubulin assembled into a short microtubule. Illustration from RCSB PDB *Molecule o the Month* by David S. Goodsell [CC-BY-4.0]. Microtubules are the target of many important toxins and drugs because of their central role in cell division. Some anticancer drugs block the rapid growth of the cancer cells by blocking the normal dynamics of microtubules. (B) Paclitaxel (shown in red) binding to tubulin. The anticancer drug paclitaxel promotes the assembly, binding so tightly that disassembly is virtually impossible (PDB: 1jff). (C) Two other molecules that block microtubule assembly, both produced by plants, are shown here: colchicine is shown in yellow and vinblastine in pink (PDB: 1Z2B).

that make up the cytoskeleton, which allows the cell to maintain its shape and size (Kabsch and Vandekerckhove, 1992). Tubulin (Figure 4) polymerize into microtubules, which is also a major component of the cytoskeleton and function in many essential cellular processes, including cell division (Goodsell, 2014; Nogales et al., 1998; Sharp et al., 2000). Tubulin-binding drugs have been developed, which inhibits microtubule dynamics essential for chromosome segregation and cell division and, therefore, can kill cancer cells (Xiao et al., 2006). Frequently used tubulin inhibitors in the treatment of cancer include paclitaxel, docetaxel, vinblastine, vincristine, and vinorelbine (Stanton et al., 2011). These small molecule drugs are also known as mitotic inhibitors because they inhibit mitotic spindle formation during cell division by interacting with tubulin (Warfield and Bouck, 1974).

Aggregators and their inhibitors. It is often argued that a protein needs a definite three dimensional structure to fulfill its function (Tsvetkov et al., 2009). However, a class of proteins is known that lacks a fixed or ordered three dimensional structure. Such a protein is called an intrinsically disordered protein (IDP) (Oldfield and Dunker, 2014). Despite their lack of stable structure, IDPs are a very large and functionally important class of proteins (Dyson and Wright, 2005). In some cases, IDPs can adopt a fixed three-dimensional structure after binding to other macromolecules. Numerous IDPs are associated with human diseases, including cancer, cardiovascular disease, amyloidoses, neurodegenerative diseases, and diabetes (Babu et al., 2011). The aggregation of the intrinsically unstructured protein α-Synuclein is thought to be responsible for Parkinson's disease. Amyloid beta aggregation is often associated with the Alzheimer's disease. Figure 5 illustrates the self-association and fibril formation for some of the important IDPs (Goodsell, 2015). Many key oncogenes have large intrinsically unstructured regions, for example p53 and BRCA1. IDPs, such as α-synuclein, tau, amyloid beta, p53, and BRCA1, are attractive targets for drugs modulating protein-protein interactions (Arkin and Wells, 2004; Metallo, 2010). Our work on the sequence complexity of amyloidogenic regions in IDPs (Das et al., 2014) and on the lysozyme aggregation (Das et al., 2013) has been previously published. Here, in the chapter 3 of this

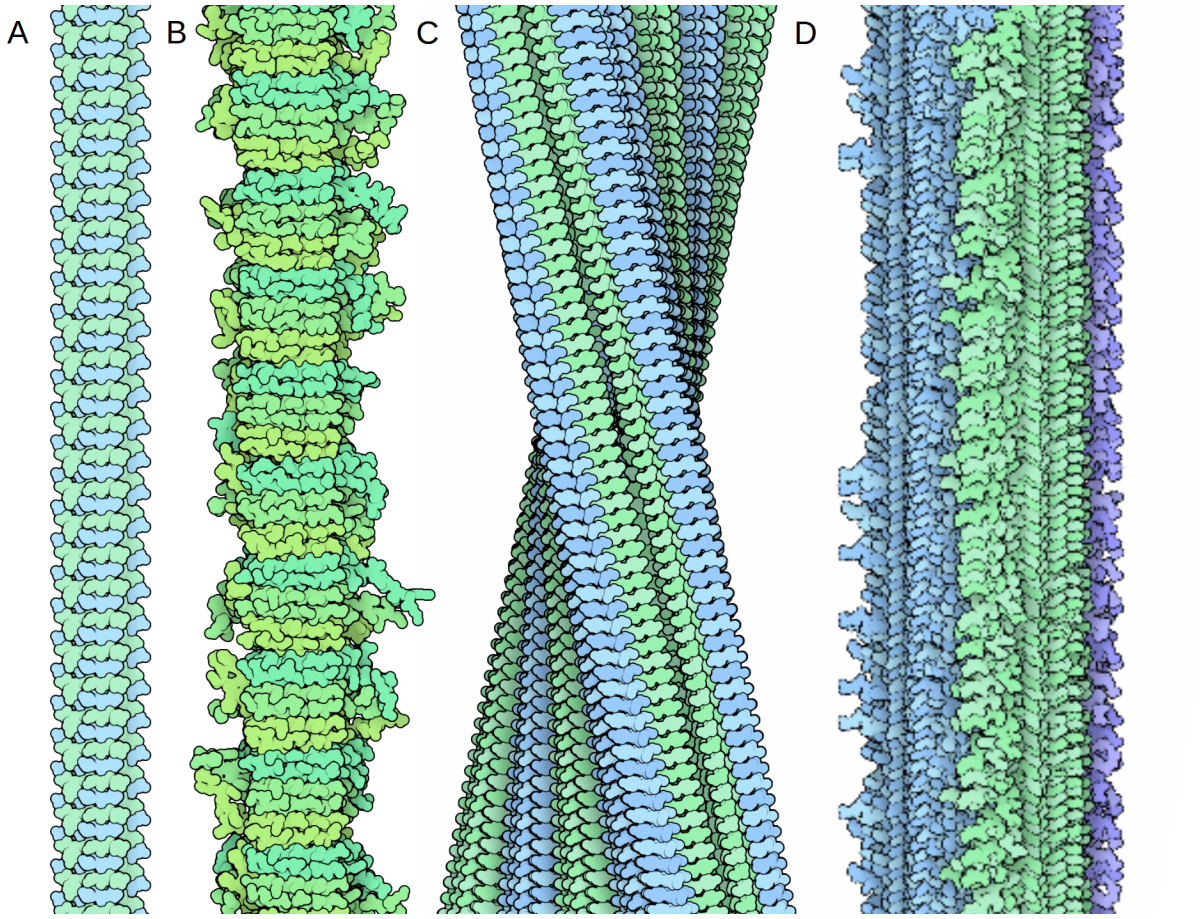


Figure 5: Amyloids. (A) Amyloid fibrils of a peptide from human prion protein (PDB: 3NHC). (B) Amyloid fibrils of a peptide from yeast prion HET-s (PDB: 2KJ3). (C) Amyloid fibrils of a peptide from transthyretin (PDB: 3ZPK). (D) Amyloid-beta fibril from a patient with Alzheimer's disease (PDB: 2M4J). The amyloids often act as a template to nucleate the misfolding and aggregation of normal proteins in the body, leading to the formation of many more amyloid fibrils. Drugs and peptidomimetics that can stop the growth of the fibrils may prevent the disease progression. Illustrations from RCSB PDB Molecule of the Month by David S. GoodSell [CC-BY-4.0].

thesis, a detailed statistical analysis on the human disordered proteome and the disordered binding regions has been given (Pal et al., 2016).

Target proteins. A biological target is anything within the body to which an endogenous ligand or a drug is directed and/or binds. Biological targets are most commonly proteins such as enzymes, ion channels, and receptors. Nucleic acids can also be a biological target; for example cisplatin, the first member of a class of platinum-containing anti-cancer drugs, binds to and cause crosslinking of DNA, which ultimately triggers cell death (Jung and Lippard, 2007; Wang and Lippard,

2005). However, in pharmaceutical research the native protein in the body whose activity is modified by a drug resulting in a desirable therapeutic effect is often referred to as a target. The completion of the human genome project in 2001 revealed that the total number of human genes is about 30,000 (Lander et al., 2001). However, the number of potential drug targets coded by the human genome may in fact be much smaller than originally speculated. Of all the currently marketed drugs with known mode of action only 266 acts through human genome derived proteins and 58 drugs acts through bacterial, viral, fungal or other pathogenic organism targets (Overington et al., 2006). An assumption that proteins related down to 50% identity show related pharmacology, expands this list to 604 genes and inclusion of all the homologues (at 35% identity) expands this number to 1,048 genes compared to ~30,000 genes in human genome (Hopkins and Groom, 2002; Overington et al., 2006). A common property of most current drug targets is that 60% of the drug targets are cell surface proteins. There, are 1,620 distinct human protein sequences that are linked directly to a genetic disease. However, only 105 of these are drug targets, corresponding to 47% of human drug targets that are directly associated with a disease (Overington et al., 2006). However, the actual number of druggable proteins could be much bigger, for which no drugs have been developed yet, for example, the disordered proteome. In the chapter 2 of this thesis we have discussed the computational approach to find allosteric druggable sites in known proteins and in the next chapter druggable binding regions in disordered proteome have been explored.

Druggable proteins. The most common drug targets of currently marketed drugs include: G protein-coupled receptors, enzymes, ion channels, nuclear hormone receptors, structural proteins such as tubulin, membrane transport proteins and nucleic acids (Figure 6). However, about 27% of the drugs target the G protein coupled receptors (Overington et al., 2006). Among the G protein coupled receptors, the most prominent targets are dopamine receptors, histamine receptors, adrenoceptors and short-peptide receptors. More than 50 FDA-approved drugs are available for each of these G protein coupled receptor groups (Overington et al., 2006). Dopamine receptors are implicated in many neurological

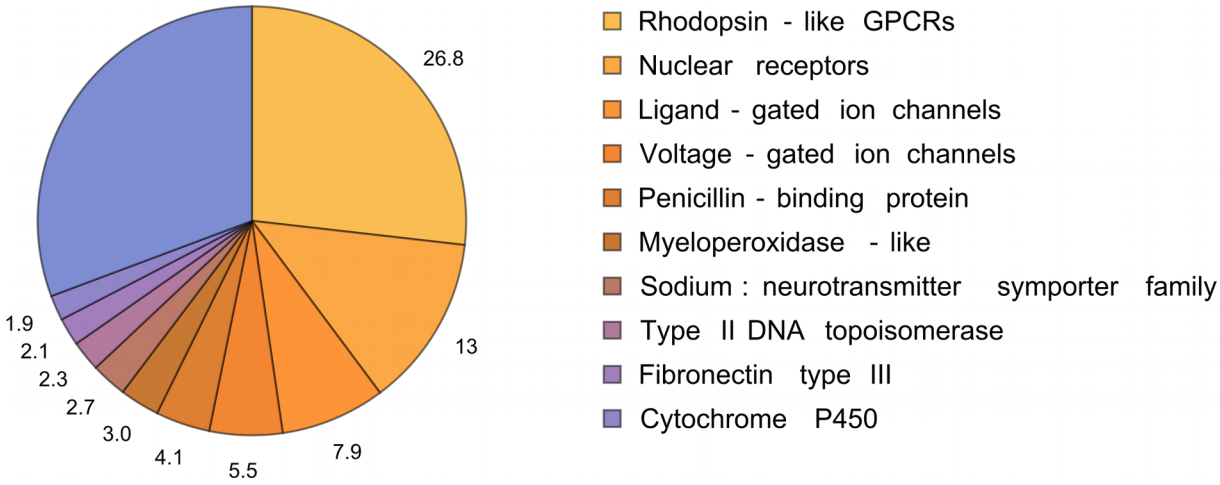


Figure 6: Target protein distribution of current drugs per drug substance. The family share as a percentage of all FDA-approved drugs is displayed for the top ten families of target proteins. Beyond the ten most commonly drugged protein families, all the other families (around 120 domain families) for which only a few drugs are available are not shown here. Together they constitute ~31% of druggable targets.

processes and the abnormality in receptor signaling leads to many neuropsychiatric disorders including Parkinson's disease and schizophrenia (Bernheimer et al., 1973). In contrast to the dopamine receptors, histamine receptors are expressed throughout the body and implicated in several functions including allergic reactions and gastric acid secretion. Antihistamines, which act on this receptor, are used as anti-allergy drugs and the H_2-antagonists, such as ranitidine and cimetidineare and used in the treatment of acid-related gastrointestinal conditions. Adrenoceptors are a class of G protein coupled receptors that are targets of adrenaline and noradrenaline. Many important and commonly prescribed medications are adrenergic antagonists, including prazosin and propanolol, which are used to treat high blood pressure, anxiety and panic disorder. Angiotensin, vasopressin, endothelin, neurokinin, nociceptin and opioids are endogenous short peptides that target a class of rhodopsin-like G protein coupled receptors, which are important targets for the peptide drugs. Other important GPCRs targets include serotonin receptors and the acetylcholine receptors. Apart from GPCRs voltage gated ion channels and membrane transporter proteins are the two class of membrane proteins for which there are more than 50 FDA-approved drugs (Overington et al., 2006). Prominent targets among the enzymes includes metalloproteases,

carboxylic ester hydrolases, phosphodiesterases, amine oxidases, cyclooxygenases, cytochrome P450s and HMG-CoA reductase. There are 10-50 drugs available for each of these group of enzymes (Overington et al., 2006). The type III nuclear receptors, which includes the androgen, oestrogen and progesterone receptors and the glucocorticoid and mineralocorticoid receptors, are also the salient drug targets having more than 50 FDA-approved drugs available in the market (Overington et al., 2006). Thyroid hormone receptors and retinoic acid receptors which belongs to the type II nuclear receptors are also one of the important class or drug targets. Among these prominent drug targets, we have worked on a metalloprotease enzyme (Rudra et al., 2012) as highlighted in the chapter 1 of this thesis. The other targets we have worked on are macrophage migration inhibitory factor (Alam et al., 2011, 2012) and EGFR kinase (Bhowmik et al., 2013), for which not many drugs are available in the market.

Types of interactions. Chemical substance when physically binds to the target proteins may invoke different responses depending on the type of interaction. The interaction between the substance and the target may be noncovalent or covalent in nature. A noncovalent binding is a relatively weak interaction between the ligand and the target where no chemical bond is formed between the two interacting molecules and hence the interaction is completely reversible. Covalent interactions are generally permanent as the ligand forms irreversible chemical bond with the target. However, a reverse reaction may readily occurs in which the bond can be broken. Depending on the nature of the ligand, two things may happen. There may be no direct change in the target protein, except that the binding of the ligand prevents other endogenous substances to bind to the target; depending on the nature of the target, this effect is referred as receptor antagonism, enzyme inhibition, or ion channel blockade. On the other hand, there may be a conformational change in the target protein induced by the ligand resulting in a change in target function; if that mimics the effect of the endogenous substance is referred to as receptor agonism or channel/enzyme activation and if the effect is opposite of the endogenous substance is called inverse agonism. These two phenomena can also be collectively called allosteric modulation. The majority of

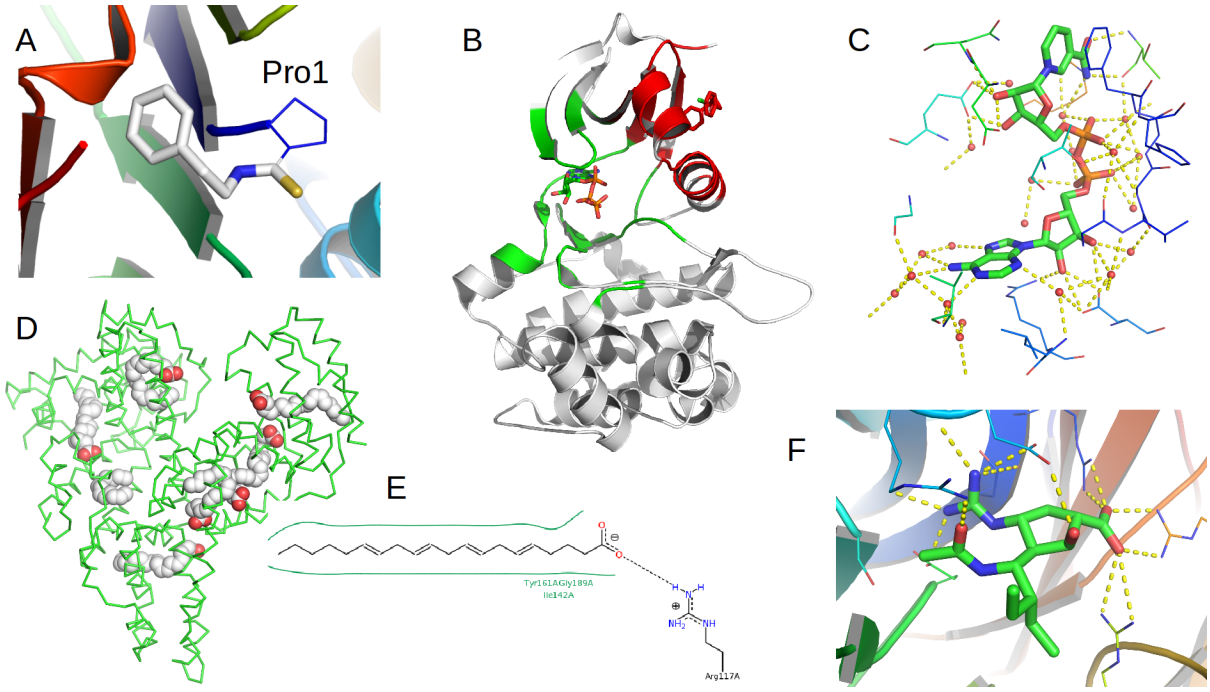


Figure 7: Different types of interactions between proteins and small molecules. (A) Covalent complex formation (suicide inhibition) between phenethylisothiocyanate and macrophage migration inhibitory factor (PDB: 3SMB). (B) Allosteric activation of phosphoinositide-dependent kinase-1 with (2Z)-5-(4-chlorophenyl)-3-phenylpent-2-enoic acid (PDB: 3HRF). ATP bound active site is colored in green and the allosteric site is colored in red. (C) Interaction through water bridges (PDB: 1LJ8). Water molecules often play an important role in the small molecule binding binding. (D) Hydrophobic interaction is another major player in protein small molecule interaction due to the entropy of water exclusion. The figure shows the binding of arachidonic acid, a long chain fatty acid with serum albumin (PDB: 1GNJ). (E) Deep hydrophobic grooves in serum albumin. Interaction diagram is generated by PoseView from PDB file 1GNJ. (F) According to the Lipinski's rule of file hydrogen bonding is one of the most important factor in protein small molecule interactions. The figure shows hydrogen bonding between influenza virus neuraminidase and small molecule BCX-1812 (PDB: 1L7F).

drug antagonists achieve their potency by competing with endogenous ligands or substrates at structurally defined binding sites on receptors (Hopkins and Groom, 2002). However, the recent developments suggested that allosteric modulators which can interact with the binding sites on the receptor molecule that are distinct from the orthosteric site could offer several advantages over antagonists, including greater selectivity (Christopoulos, 2002). Our research as described in the chapter 2 of this thesis detailed the computational method for allostery and drugability prediction in structured proteins. In chapter 3, we further explored all possible

binding sites on disordered proteins. Figure 7 illustrated some of the important modes of interactions between the protein and small molecules, including hydrophobic and hydrogen bonding contributions. Here, in the chapter 4 of this thesis, we have explored the micro-environment of the hydrophobic grooves on target proteins with a novel fluorophore. Chapter 5, on the other hand, explores the hydrogen bond strength determination between a donor and acceptor using isotope labeling and nuclear magnetic resonance spectroscopic methods.

Biologically important small molecules and peptides. The different class of target proteins have been discussed above. Most of these target proteins function via binding to the endogenous ligands, which may be small molecules or peptides or even proteins such as the growth factors or cytokines. The endogenous ligands, e.g., substances for enzymes and transporters or agonists for receptors provide the templates and serves as the lead molecule for rational drug design (Silverman and Holladay, 2014). Identification of suitable lead compounds provide starting point for lead optimization during which the leads are modified to achieve requisite potency and selectivity as well as the ADME (absorption, distribution, metabolism, and excretion) properties (Oprea et al., 2001). Figure 8 shows some important biological small molecules and peptides and the drugs developed on some of those leads.

Neurotransmitters. Neurotransmitters are an important class of small molecules that served as lead compounds for the discovery of many drugs (Silverman and Holladay, 2014). Some common small molecule neurotransmitters are acetylcholine, noradrenaline (norepinephrine), dopamine, serotonin and gamma aminobutyric acid (GABA). Dopamine is the endogenous ligand for dopamine receptors. The drug rotigotine, which is used in the treatment of Parkinson's disease was developed by modifying dopamine (Chen et al., 2009). The antimigraine drug frovatriptan and the antihypertensive drug nebivolol were developed by chemical modification of serotonin and norepinephrine, respectively (Broeders et al., 2000; Busto et al., 2013). Cevimeline, which is used in the treatment of dry mouth is an acetylcholine analogue (Petrone et al., 2002). Some neurotransmitters

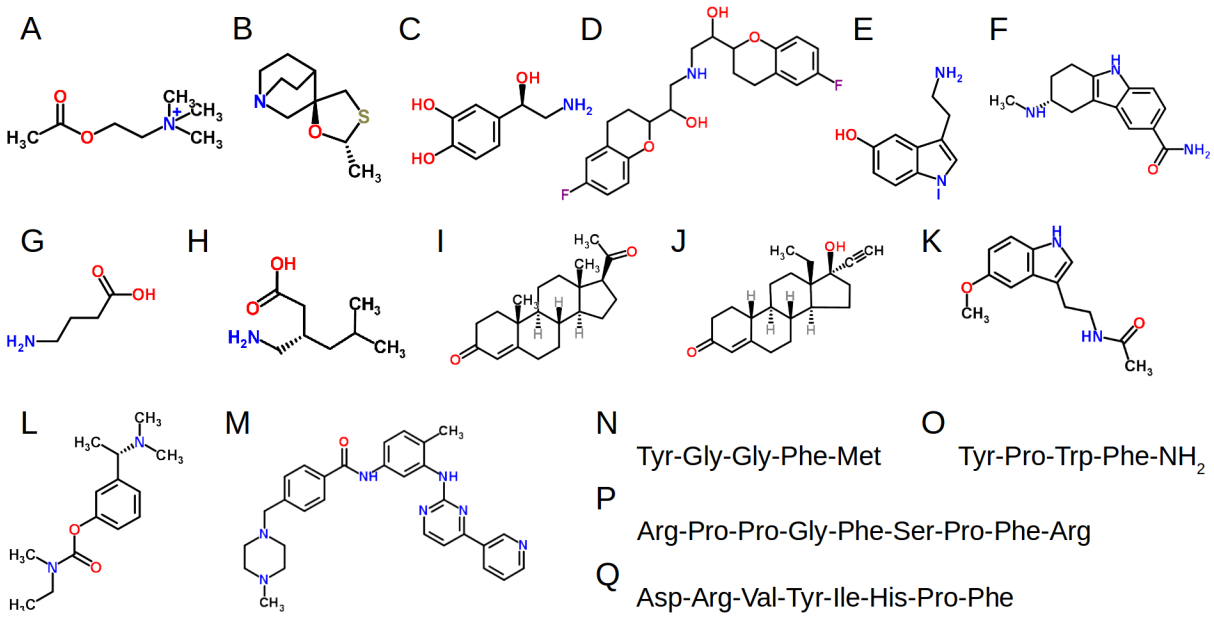


Figure 8: Some biologically important small molecules and drugs developed on these leads. (A) Neurotransmitter acetylcholine. (B) Cevimeline, a muscarinic acetylcholine receptor agonist. (C) Neurotransmitter norepinephrine (noradrenalin). (D) Nebivolol, a highly selective blocker of β1 adrenergic receptors. (E) A monoamine neurotransmitter, serotonin (5-hydroxytryptamine). (F) Frovatriptan, a 5- hydroxytryptamine receptor agonist used for the treatment of migraine headaches. (G) The chief inhibitory neurotransmitter in the mammalian central nervous system, γ-Aminobutyric acid (GABA). (H) Pregabalin (β-isobutyl-γ-aminobutyric acid), a GABA analogue used to treat epilepsy, neuropathic pain, fibromyalgia and generalized anxiety disorder. (I) An endogenous steroid hormone progesterone. (J) Levonergestrel, a more potent synthetic analogue of progesterone used in birth control medication. (K) Hormone of the pineal gland, melatonin. (L) Rivastigmine, a substrate analogue for acetylcholinesterase. It is used for the treatment of mild to moderate dementia of the Alzheimer's and Parkinson's disease. (M) Imatinib binds close to the ATP binding site of bcr-abl kinase and inhibits the enzyme activity semi-competitively. (N) Endogenous opioid peptide Met-enkephalin. (O) Endogenous opioid peptide endomorphin-1. (P) Bradykinin, an inflammatory mediator peptide. (Q) Angiotensin II, a peptide hormone that causes vasoconstriction and increases blood pressure.

such as adrenaline (epinephrine), noradrenaline (norepinephrine) and dopamine are themselves used as drugs.

Hormones. Hormones are another important class of substances that have served as lead compounds for drug discovery (Silverman and Holladay, 2014). Like neurotransmitters, hormones are secreted from cells and interact with receptors

on other cells. However, unlike the neurotransmitters, the site of action of the hormone is not adjacent to the site of release. Hormones travels through the bloodstream to their site of action. Steroids are one of the important classes of hormones. The endogenous steroid hormones such as progesterone and estrogen show weak and short lasting effects and, therefore, chemically modified more potent synthetic progesterone (levonorgestrel) and estrogen (17α-ethynyl-estradiol) are used in the formulations of oral contraceptives (Folmar et al., 2000; Park, 2012). Melatonin is another important hormone secreted from pineal glands and is involved in the synchronization of the circadian rhythms of physiological functions. The hormone is itself used as a sleep aid and in the treatment of some sleep disorders. A study on the alternative target for melatonin has been highlighted in the chapter 1 of this thesis, where we have shown that the hormone melatonin binds with the gelatinase B with high affinity (Rudra et al., 2012).

Peptides. We discussed about the small molecule neurotransmitters and hormones. However, peptides constitute another broad class of hormones and neurotransmitters (neuropeptides). Peptides, like proteins consist of a sequence of amino acid residues, but are smaller than proteins. The length of peptide may vary from two amino acids to approximately 50 amino acids. Small peptide hormones include thyrotropin-releasing hormone, vasopressin, angiotensin and the natriuretic peptides (hormones of the heart). Some peptide hormones such as somatostatin and cholecystokinins also acts as neurotransmitters. Neuropeptides are used by neurons to communicate with each other. Many neuropeptides are co-released with other small-molecule neurotransmitters. Important neuropeptides include tachykinins and opioid peptides. Tachykinins are from ten to twelve residues long and they excite neurons, evoke behavioral responses, are potent vasodilators, and contract smooth muscles of gut tissue. The search for endogenous ligands for morphine receptors led to the discovery of opioid peptides such as enkephalins and endorphins, which alleviates pain. Unlike these natural analgesics, nociceptin is a neuropeptide which increased sensations of pain. Endothelins are another class of peptides secreted by endothelial cells and involved in vascular homeostasis. An honorable mention of the important

endogenous peptides is amyloid-β peptides which are 40-42 amino acid long peptides fromed by digestion of amyloid precursor protein. The 42 amino acid long amyloid-β is toxic as it aggregates and forms senile plaques associated with Alzheimer's disease (Cheng et al., 2013). Considerable effort has been devoted to the goal of using natural peptides for the discovery of derivatives with improved properties (Silverman and Holladay, 2014). Most peptides have low stability in plasma due to the presence of peptidases, which hydrolyzes the peptides into smaller peptides or constituent amino acids (Adessi and Soto, 2002). Moreover, peptides usually cannot be delivered orally because they get digested in the gut. However, cross-linking with disulfide bonds may confer enzymatic stability to the peptide drugs. Our study on anticancer peptide incorporating unusual D-amino acids has been previously published (Banerji et al., 2012). Here in the chapter 1 of this thesis, interaction of benzyl protected cysteine-based dipeptides with serum proteins has been highlighted (Banerji et al., 2013a). These peptides formed unbranched nanotubes in solution (Banerji et al., 2013b)and also showed anti-cancer activity (Banerji et al., 2013a).

Substrates. The discussion of the biologically important endogenous small molecules and peptide ligands so far has focused on leads for dugs designed to interact with receptor targets. Endogenous ligands for other types of drug targets, including transporters and enzymes, have also served as valuable starting points for drugs (Silverman and Holladay, 2014). As mentioned previously, transporters are proteins that help transport substances such as neurotransmitters, glucose and ions across cell membranes. Unlike acetylcholine, which is hydrolyzed in the synaptic cleft, some neurotransmitters such as dopamine, serotonin and norepinephrine are recycled. Inhibitors of the reuptake transporters for these neurotransmitters comprise important classes of antidepressant drugs (Zhou et al., 2007). The leads for many of these reuptake inhibitors were the transporter ligands themselves, that is, dopamine, serotonin or norepinephrine. Transporters of glucose have recently been targeted for the treatment of type II diabetes (Idris and Donnelly, 2009). Enzyme substrates and the transition sate analogues are an important source of leads for the design of enzyme inhibitors (De Clercq, 2002; Dreyer et al., 1989; Schramm, 2013). Rivastigmine which is an inhibitor of

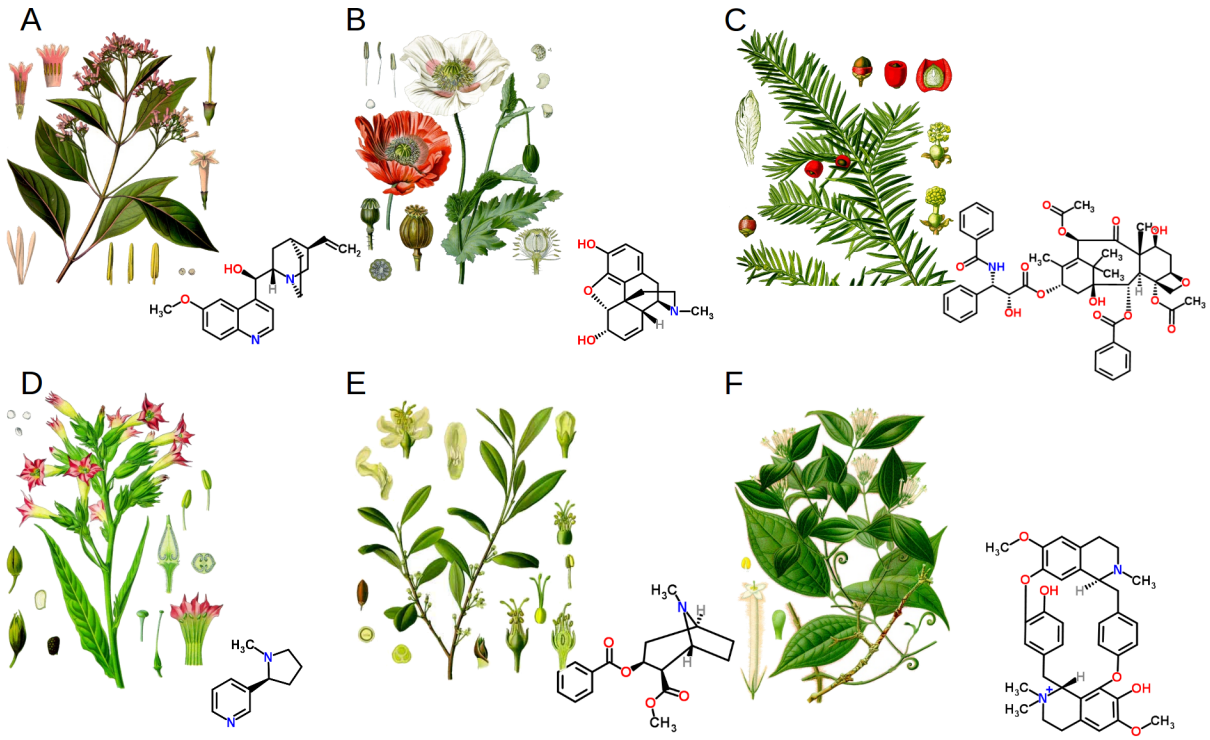


Figure 9: Drugs obtained from plant sources. (A) The bark of *Cinchona officinalis* L. is a rich source of antimalarial drug quinine. (B) Morphine is obtained from opium poppy (*Papaver somniferum* L.). (C) Chemotherapeutic drug paclitaxel is harvested from the leaves of European yew (*Taxus baccata* L.). (D) Acetyl choline receptor agonist nicotine is found in tobacco (*Nicotiana tabacum* L.). (E) *Erythroxylum coca* Lam. produces cocaine, a reuptake inhibitor of serotonin, norepinephrine and dopamine. (F) D-tubocurarine, the main toxin of arrow poison is obtained from *Strychnos toxifera* M. R. Schomb. ex Benth. Illustrations of the plants are in public domain.

acetylcholine inhibitor used for the treatment of dementia was developed from the acetylcholine lead (Polinsky, 1998). ATP (adenosine triphosphate), which plays a central role in the intracellular energy transfer and is a substrate to more than 500 human protein kinases, served as a lead for development of many kinase inhibitors including imatinib (Hantschel et al., 2012; Zhang et al., 2010), which is used for the treatment of chronic myeloid leukemia.

Natural products. Nature is also a rich source of drug precursors and in some cases of actual drugs (Koehn and Carter, 2005; Newman and Cragg, 2007). Small molecules obtained from plants, marine organism, bacteria and fungi constitutes an impressive arsenal of natural products that have implications in a plethora of biological activities. Penicillin derived from fungi is a historically important

antibacterial agent. Other important antibiotics derived from bacteria includes tetracycline the polymyxins, and the rifamycins. Plants are a major source of complex and structurally diverse chemical compounds. Clinically useful examples include the anticancer agent paclitaxel (Sandler et al., 2006), the antimalarial agent artemisinin (White, 2008), and the acetylcholinesterase inhibitor galantamine (Heinrich and Lee Teoh, 2004). Other plant-derived drugs, used medicinally and/or recreationally include morphine, cocaine, quinine, tubocurarine, muscarine, and nicotine. Figure 9 shows some natural products along with their plant sources. Animals, in particular, venomous animals also represent a source of bioactive natural products. Teprotide, a peptide isolated from the venom of the Brazilian pit viper was a lead in the development of the antihypertensive agents cilazapril and captopril (Clozel, 1991). Analgesic agent ziconotide is the synthetic form of the neurotoxic peptide ω-conotoxin isolated from the marine cone snail (Duggan and Tuck, 2015). Trabectedin derived from the marine tunicate is used to treat metastatic soft tissue sarcoma (Petek et al., 2015). Other natural products derived from marine animals include the antitumour agents discodermolide (Dall'Acqua, 2014), and the bryostatins (Kollár et al., 2014). A large number of FDA-approved drugs have been either directly derived from or inspired by natural products (Patridge et al., 2015). Natural products are a continuing source for novel drug leads. In the chapter 1 of this thesis, a work on the discovery of a macrophage migration inhibitory factor inhibitor from *Azadirachta indica* (Alam et al., 2012) has been highlighted.

Drugs and probes. A drug is a chemical substance that has known biological effects. Most drugs are small molecules, although some drugs can be proteins (Overington et al., 2006). Small molecule drugs generally follow the Lipinski's rule of five, which describes the molecular properties important for a drug's pharmacokinetics in the human body, including their absorption, distribution, metabolism, and excretion (Lipinski et al., 1997). The rule sates that for an orally active drug the molecular mass should be less than 500, octanol-water partition coefficient (logP) should not be greater than 5 and there should be no more than 5 hydrogen bond donors and 10 hydrogen bond acceptors. However, not all

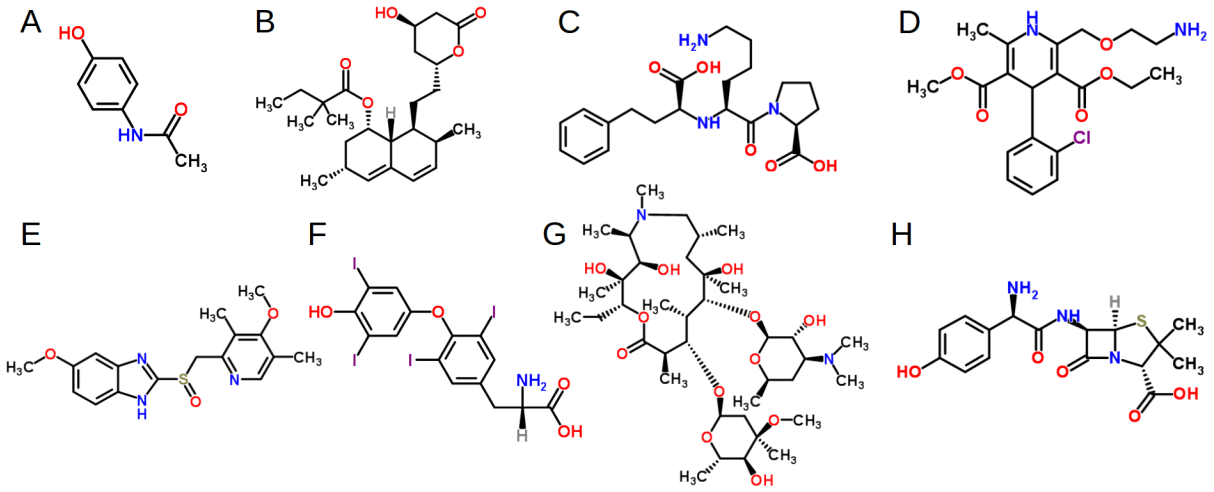


Figure 10: Most commonly prescribed drugs. (A) Paracetamol, a non-opioids and non-steroidal anti-inflammatory medicine used for pain and palliative care. (B) Simvastatin, a lipid lowering medication used to decrease the risk of heart problems in those at high risk. (C) Lisinopril, an angiotensin-converting enzyme inhibitor, used primarily in treatment of high blood pressure, heart failure, and after heart attacks. (D) Amlodipine, a medication used to lower blood pressure and prevent chest pain. (E) Omeprazole, an irreversible proton pump inhibitor, used to treatgastroesophageal reflux disease and ulcer. (F) Levothyroxine, a synthetic thyroid hormone used to treat thyroid hormone deficiency. (G) Azithromycin, a broad spectrum antibiotic. (H) Amoxicillin, another broad spectrum antibiotic.

molecules following this rule qualify to be drugs. Most commonly prescribed drugs include paracetamol to reduce pain and fever, cholesterol lowering drug simvastatin, blood pressure medication lisinopril and amlodipine, antacid drug omeprazole, thyroid drug levothyroxine and the antibiotics azithromycin and amoxicillin. Figure 10 shows the chemical structures of some of the most commonly prescribed drugs.

Existing drug space. Database searches produce an excess of 21,000 drug products; however, when duplicate active ingredients, salt forms, supplements, vitamins, imaging agents, and so on are removed, this number is reduced to only 1,357 unique drugs, of which 1,204 are small-molecule drugs and the rest are biological drugs (Overington et al., 2006). 803 drugs out of the 1,204 small-molecule drugs, can be administered orally, 421 can be administration as injection or infusion; 275 are administered using other routes including buccal, rectal and

inhalational. At least 16% of the small-molecule drugs are prodrugs and 70% of orally dosed drugs pass the rule-of-five test, whereas 20% of orally dosed drugs fail at least one of the rule-of-five parameters (Overington et al., 2006).

Drug repurposing. Although a drug is developed to target a particular protein, small molecule drugs may have alternate targets. Drug development is a lengthly process; translation of a promising molecule into an approved drug often takes more than 14 years and cost over one billion dollars. However, discovering new uses for approved drugs provide the quickest possible transition from bench to bedside (Ashburn and Thor, 2004; Mullard, 2012). Examples of drug repurposing includes metformin (Pryor and Cabreiro, 2015), sildenafil and thalidomide. Sildenafil, originally developed as antihypertensive drug and failed, is now widely used to treat erectile dysfunction. Once over-the-counter drug thalidomide, which was banned due to birth defects, has been revived for the use in leprosy and multiple myeloma (Bartlett et al., 2004). Identifying and developing new uses for existing drugs is an emerging field and we have studied the repositioning of omeprazole as an allosteric modulator for gelatinase A (not shown). We have also found a new target for melatonin. Melatonin, which is used as a supplement to treat sleep disorders, can also cure stress induced gastric ulcer in mice via inhibition of gelatinase B (Rudra et al., 2012). Interactions of melatonin with gelatinase B has been highlighted in the chapter 1 of this thesis.

Chemical probes. Small molecules that leads in the development of new therapeutic agents may also be used as research tools to probe biological function. Chemical biologists frequently aim to create small-molecule probes that interact with a specific protein in vitro in order to explore the role of the protein in a broader biological context (Frye, 2010). Some can inhibit a specific function of a multifunctional protein or disrupt protein–protein interactions. However, drug leads and probes differ in important ways. Dyes, for example, are used as probes to study the biophysical properties of proteins (Woestenenk et al., 2003). The dye thioflavin T is widely used to visualize and quantify the presence of misfolded protein aggregates called amyloid (Das et al., 2013), both in vitro and in vivo (Das et

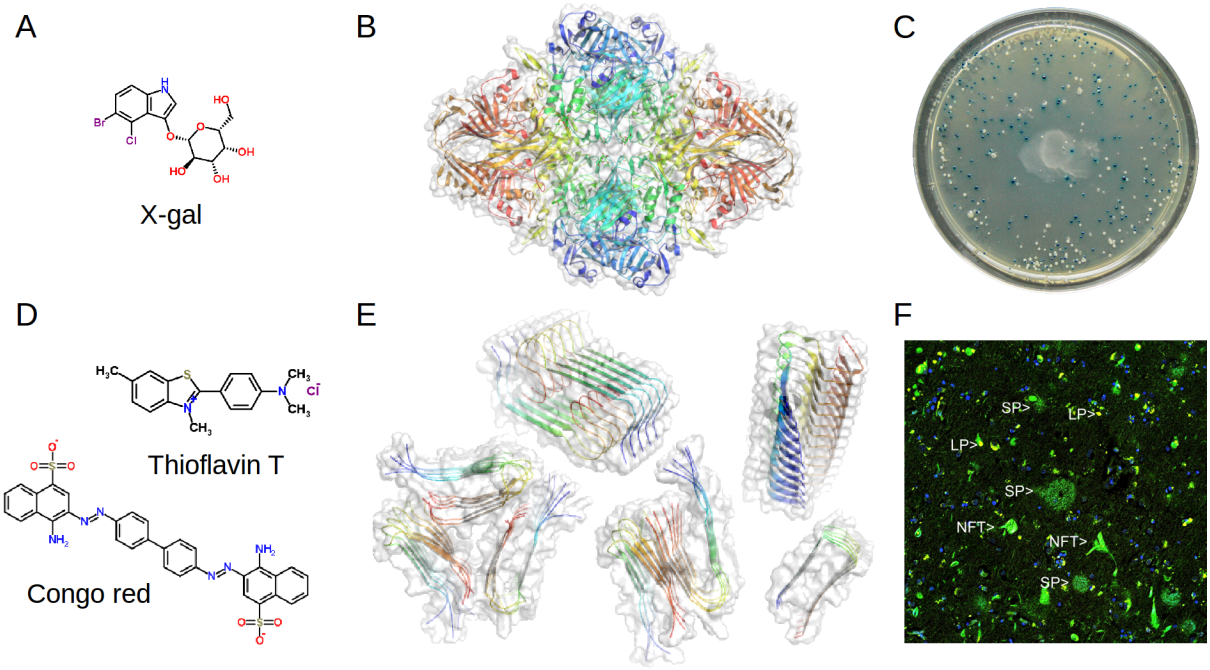


Figure 11: Chemical probes and their implications in biology. (A) X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) is a substrate based probe widely used in molecular biology specifically in vector-based molecular cloning experiments for the detection of recombinant bacteria. (B) E. coli β-galactosidase (lacZ), the enzyme that can hydrolyze x-gal producing insoluble blue compound, thus, indicating unsuccessful cloning. PDB ID: 3DYO. (C) An LB agar plate showing the result of a blue white screen by Stefan Walkowski [CC BY-SA 4.0-3.0-2.5-2.0-1.0]. (D) Thioflavin T and Congo red are fluorescent probes used as sensitive diagnosis tools for amyloidosis. Congo red is used in traditional histological birefringence test amyloidosis as well. (E) X-ray crystallographic structures of amyloid-β aggregates. PDB IDs: 2BEG, 2LMN, 2LMP, 2MXU, 2MVX. (F) Alzheimer's hippocampus section stained with Thioflavin S, which binds to both senile plaques (SP) and neurofibrillary tangles (NFT), the two hallmark lesions of Alzheimer's disease. Autofluorescent lipofuscin granules (LP) can also be seen in this image (Source: http://encorbio.com/monoclonal/MCA-AB9.html. EnCor Biotechnology Inc. Retrived January 14, 2016) [CC BY-SA 3.0].

al., 2013; Krebs et al., 2005). X-gal is another probe much used in molecular biology to test for the presence of an enzyme, β-galactosidase. Apart from the protein aggregation or enzyme activity, fluorescent molecules are also used to probe the microenvironment such as the polarity of protein binding sites. Fluorescein and rhodamine are two small molecules, which are toxic for therapeutic use; however, they are widely used as fluorescent tracers for many applications. Figure 11 highlights some important chemical probes routinely used in molecular biology and structural biology. Figure also shows the target proteins and the representative

assay results. In the chapter 4 of this thesis, we reported a novel naphthalene based fluorescent probe and its behavior in the hydrophobic micro-environment of the protein binding site (Pal et al., 2015).

Methods to study the protein small molecule interactions. In rational drug development millions of potential drug molecules are tested for a protein target, which requires fast, robust and reliable high throughput screening methods. The importance of protein small molecule interactions in drug discovery and protein metabolite interactions in biology has driven the development of new methods that rely heavily on the integration of both the theoretical and experimental chemistry (McFedries et al., 2013).

Docking and dynamics. Molecular docking and dynamics are computational methods that provide ample information about the protein small molecule interactions such as the binding free energy, binding site, interacting residues, types of interactions and how it alters the protein structure (Meng et al., 2011; Morris and Lim-Wilby, 2008). However, these methods can only be applied to the proteins for which the structural information is available. Docking is the basis for virtual screening, a high throughput screening for drug leads from a library of millions of drug like compounds. These methods have been used by academics and the pharmaceutical companies in the hunt for drug leads and have become an indispensable part of structure based rational drug design. Over the years, several docking algorithms have been developed by different academic groups and benchmarked against the known protein ligand complexes established by more direct methods such as X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy (Friesner et al., 2004; Grosdidier et al., 2011, 2011; Mashiach et al., 2008; Morris et al., 2009; Schneidman-Duhovny et al., 2005). These methods are fast, cost effective, robust, and reliable. Docking and dynamics depends on the assumptions of molecular mechanics and, therefore, does not allow probing covalent interactions with the proteins. However, this obstacle can be overcome by applying quantum mechanical or hybrid quantum/molecular mechanical calculations. The quantum mechanical or hybrid quantum/molecular mechanical

studies along with the docking also allows studying the mechanism of action of an enzyme in an enzymatic reaction (Tao et al., 2009). Molecular docking, dynamics and quantum mechanical studies has been extensively used in most of the studies described in subsequent chapters (Alam et al., 2011, 2012; Banerji et al., 2013a; Bhowmik et al., 2013; Pal et al., 2015; Rudra et al., 2012).

Ultraviolet-visible spectroscopy. Spectroscopy is a widely used biophysical method in the study of protein small molecule interactions. Proteins generally have two special aromatic amino acids, tyrosine and tryptophan, which absorbs ultraviolet (UV) light and fluoresces as well when shined with UV light. Besides, the backbone of a protein arranges in different secondary structures such as helix, extended or coils, which lead to the differential absorption of circularly polarized far-UV light producing circular dichroism (CD) signals (Woody, 1996). These spectral properties of protein are used extensively to study their interactions with small molecules (Gray, 1996; Lakowicz, 2006a), which is photoinactive or whose photophysical properties does not interfere with that of protein's. A small molecule may also be photoactive depending upon its conjugated system and if it is chiral, it also produces CD signal. Some achiral compounds may give CD signal as well upon binding to a protein (induced CD) (Hatano, 1986). In such cases, the spectral properties of the small molecule may also be utilized to study kinetics and thermodynamics of its binding with the target protein (Lakowicz, 2006a). Fluorescent molecules may show anisotropy of fluorescence when bound to a protein, which is again a widely used property to measure the binding constants and kinetics of interaction (Lakowicz, 2006b). Overlap of the molecule's absorption with protein fluorescence may allow resonance energy transfer from the protein to the interacting small molecule (Forster resonance energy transfer; FRET) (Lakowicz, 2006c). FRET as well as anisotropy can also elucidate rotational diffusion of the ligand and its proximity to the protein. Some fluorescent molecules when binds to the protein, depending on the binding site hydrophobicity, polarity etc., may show increased quantum yield and/or higher or lower energy emissions (blue and red shifts, respectively). Such properties of the fluorescent compound can be used to probe the microenvironment of the bindng site of the protein as explored in the

chapter 4 of this thesis. Other techniques which make use of the fluorescence properties of the proteins ligands to explore different biophysical aspects of their interaction include TCSPC (time correlated single photon count), FCS (fluorescence correlation spectroscopy), and laser flash photolysis (Lakowicz, 2006d). UV-visible spectroscopic methods were also extensively used in our previously published works (Banerjee et al., 2012; Banerji et al., 2013a, 2014; Ray et al., 2012) and in the study of protein binding site micro-environment as described in the chapter 4 of this thesis (Pal et al., 2015).

Nuclear magnetic resonance spectroscopy. One of the most important tools for evaluating the protein small molecule interactions has been high-field solution state nuclear magnetic resonance (NMR) spectroscopy because of its unique ability to examine even weak protein ligand interactions at high resolution (Pellecchia et al., 2002, 2008). NMR can be used to evaluate the structural, thermodynamic and kinetic aspects of a binding reaction (Skinner and Laurence, 2008). The protein observed methods include heteronuclear single quantum correlation (HSQC) experiments, which provides an overall map or fingerprint of the protein target and serves as the basis for assessing ligand binding. ^{1}H-^{15}N HSQC detects changes in the backbone amide bonds upon addition of ligand, whereas the ^{1}H-^{13}C HSQC detects changes in aliphatic and aromatic chemical shifts of the side chains. Ligand observation experiments, which includes exchange transferred nuclear Overhauser effect, saturation transfer difference, relaxation and diffusion experiments offer a useful alternative to protein target detection for larger proteins. The use of simple one-dimensional spectra and ability to screen mixtures, makes it a prime technique for high throughput screenings (Dalvit et al., 2002). Ligand examination has the technical advantage of requiring a smaller amount of unlabeled protein and may be used to identify the functional groups of the ligand involved in the interaction. Moreover, with isotope labeling finer details of the interactions such as strength of hydrogen bondings(Maiti et al., 2006), proton transfer, transition state intermediates can be studied with NMR (Maiti et al., 2006). Distance constrains obtained from NMR and FRET experiments can be used in docking to develop better models for the protein ligand interactions (van Dijk et al., 2005; Sturlese et

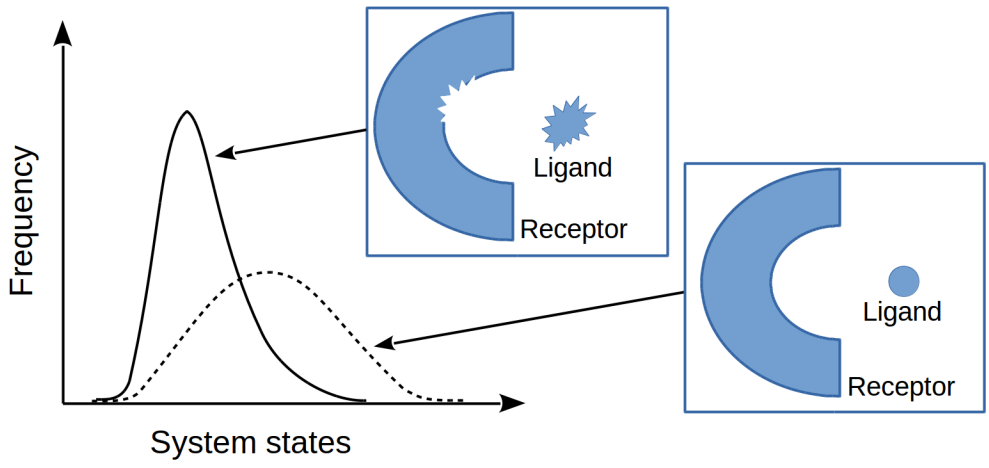
al., 2015). NMR methods has been used to study the hydrogen bonding strength in model donor acceptor complexes (Pal et al., 2014) as described in the chapter 5 of this thesis.

Other methods. X-ray crystallography is the most direct method to study the protein ligand interaction. Protein crystallography uses X-rays to reveal the structure of proteins. It used to be a laborious process, sometimes taking more than a year to determine just one structure. However, major advances in experimental methods and software have made it possible to determine the structure of a protein–ligand complex in less than a day. The high-throughput crystallography is becoming a great asset to the drug discovery process, providing unprecedented insight into the interaction of drug candidates with protein targets (Blundell et al., 2002; O'Reilly et al., 2006). There are several other methods to study the protein small molecule interactions, which includes isothermal titration calorimetry, Fourier transform infrared (FTIR) spectroscopy, Raman spectroscopy, mass spectrometry and enzyme assays. Isothermal titration calorimetry is a biophysical method that depends on the heat generation or heat absorption due to the molecular interactions. All the biochemical reactions are either exothermic or endothermic in nature and by monitoring the heat change due to interactions, thermodynamic and kinetic parameters of a reaction can be obtained with high precision. Besides, the recent advances in infrared and Raman spectroscopy techniques rendered them as valuable tools to monitor the dynamics and exact molecular details of protein ligand interactions (Kötting and Gerwert, 2013). FTIR and Raman spectroscopy yields similar, but complementary, information about the vibrational modes in the system. On the other hand, enzyme assays are vital for the study of enzyme kinetics and enzyme inhibition and are routinely used for high throughput screenings by the pharmaceutical companies (Goddard and Reymond, 2004). Design of the assay depends on the nature of the target. Generally, the catalytic activity is detected using labeled substrates or indirect sensor systems that produce a detectable spectroscopic signal upon reaction. Further, in enzyme assays, microarray formats have been devised to increase throughput. However, these methods has not been used in the studies described in this thesis.

SCOPE OF THE THESIS

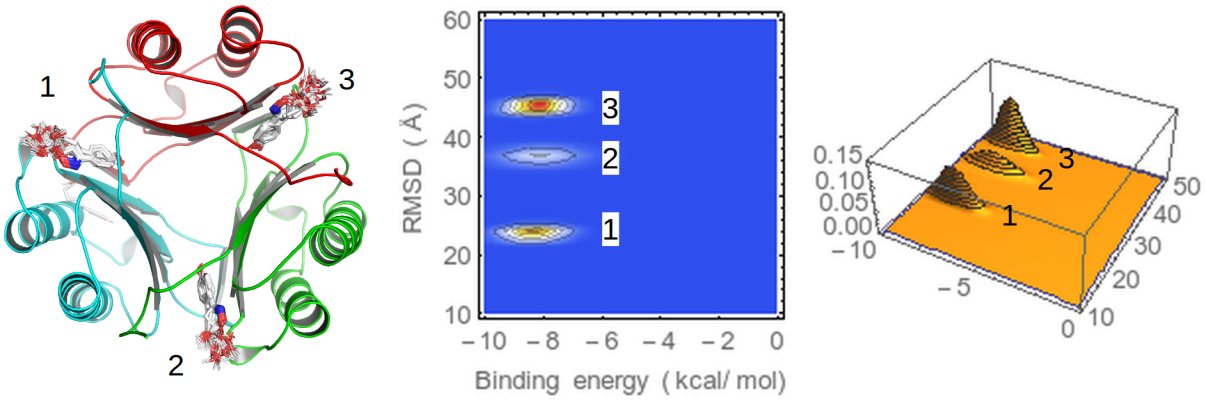
"What is that through which, if it is known,
everything else becomes known?"
Mundaka Upanishad, 1.1.3

This thesis explores the ligand protein interactions to understand basic biology and its implications in rational drug development. What would it look like if the interaction of a ligand with a target protein were completely random? How the complementarity of a ligand with its binding site on the target protein makes specific interactions possible and how can this specificity be detected in a computational analysis? How can we find the allosteric drug leads for a target protein? What are the characteristics of binding regions in an intrinsically disordered protein? How the environment that a ligand experiences upon binding to the protein differs from that of the bulk solvent? And how the specific interactions such as the strength of a hydrogen bond can be probed to better understand the ligand protein interactions? These are the questions that has been addressed in this work.



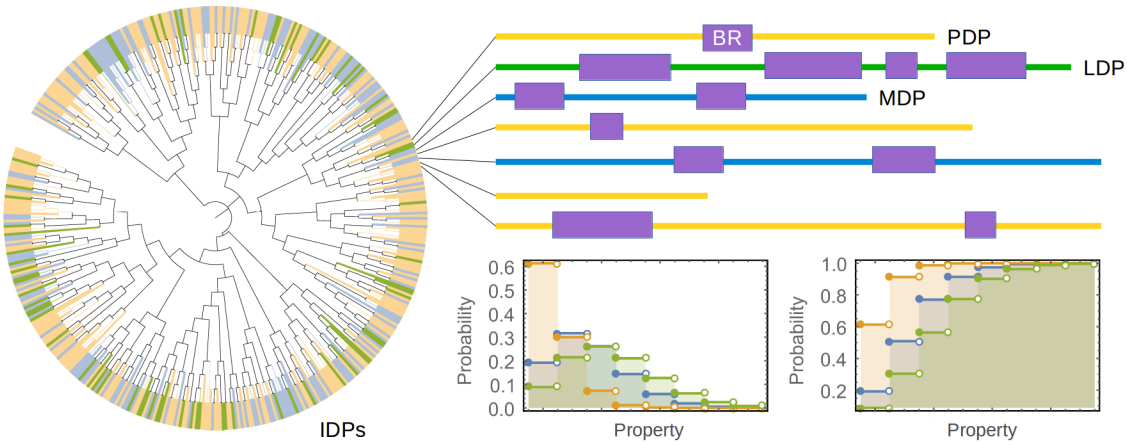
The first chapter of this thesis describes the quest to develop an improved way of finding out the potential inhibitors of a target protein from a library of small molecules. Some basic concepts abstracted from statistical mechanics, namely,

ensemble analysis, was implemented to study the specificity under the assumption of the randomness of interaction. This computational approach is based on the existing molecular docking techniques. Molecular docking, being a fast and robust technique, is widely used in the field of drug design and development. Still, the results are greeted with much skepticism. We often encounter questions regarding reliability of the result and its reproducibility. Being a stochastic technique the outcome is not exactly reproducible. That brought up the necessity of rigorous statistical analysis of the outcome. In this chapter, the ensemble analysis methods has been discussed to extract the underlying specificity information in the molecular docking simulation. It has been shown here that interactions driven by nonspecific hydrophobic forces produce a normal distribution pattern of the docking poses derived using genetic algorithm. Whereas, if the ligand protein binding is driven by some specific interactions, a positive skewness is observed in the distribution of docking poses. The skewness in the distribution, which suggests a bias towards a particular low energy conformation, improves the reliability of docking simulation many folds. How this parameter can be used as a filter in virtual screening has been discussed in this chapter.



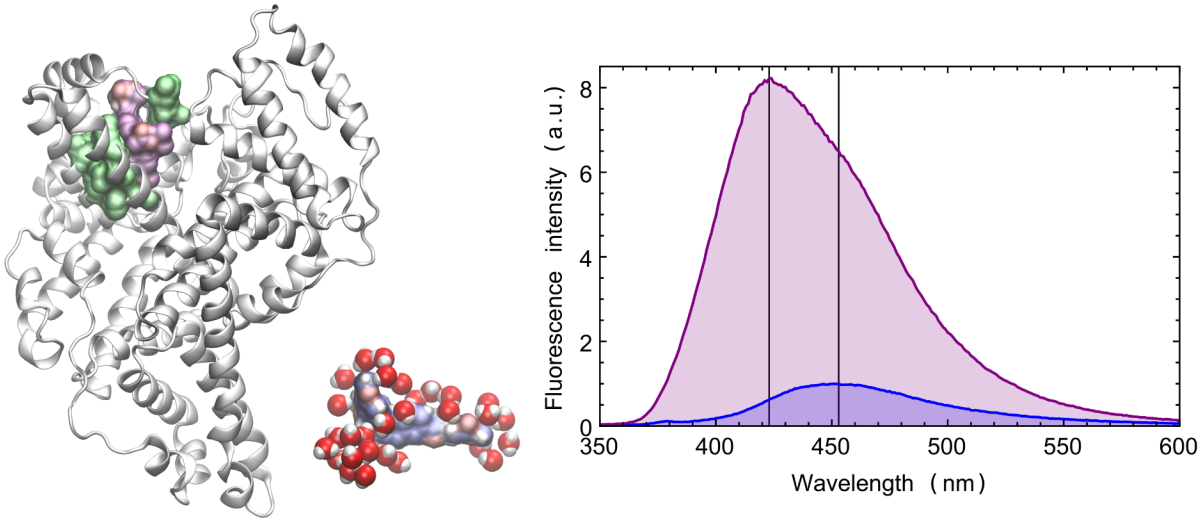
The next chapter shows a cleaver way of segregating the different type of inhibitors such as the competitive and allosteric inhibitors from a single virtual screening experiment. Allostery is a direct and efficient way of regulating biomacromolecule function. The allosteric modulators can fine-tune protein mechanics. Besides, allosteric sites are evolutionarily less conserved/more diverse

even in very similarly related proteins, thus, provides high degree of specificity in targeting a particular protein. Therefore, targeting of allosteric sites is gaining attention as a strategy in drug design. However, the experimental approaches provide a limited degree of characterization of new allosteric sites. Computational approaches are useful to analyze and select potential allosteric sites for drug discovery. In this chapter, the use of molecular docking, which has become an integral part of the drug discovery process, has been discussed to predict the druggability of novel allosteric sites as well as the active site on target proteins in a single virtual screening setup. Genetic algorithm was used in docking and the whole protein was placed in the search space. For each ligand in the library of small molecules, multiple run of the genetic algorithm populated all the druggable sites in the target protein, which was then translated into two dimensional density maps (patterns). High density clusters were observed for lead like molecules in these pattern diagrams. Each cluster in such a pattern diagram indicated a plausible binding site and the density gave its druggability score in terms of weighted probabilities. The patterns were filtered to find the leads for each of the druggable sites on the target. Such, novel pattern based analysis of the clusters provides a way to probe novel druggable sites on a target protein in a much simpler setup. This structure based analysis method might help researchers to develop allosteric modulators and to identify novel target sites on drug resistant proteins.



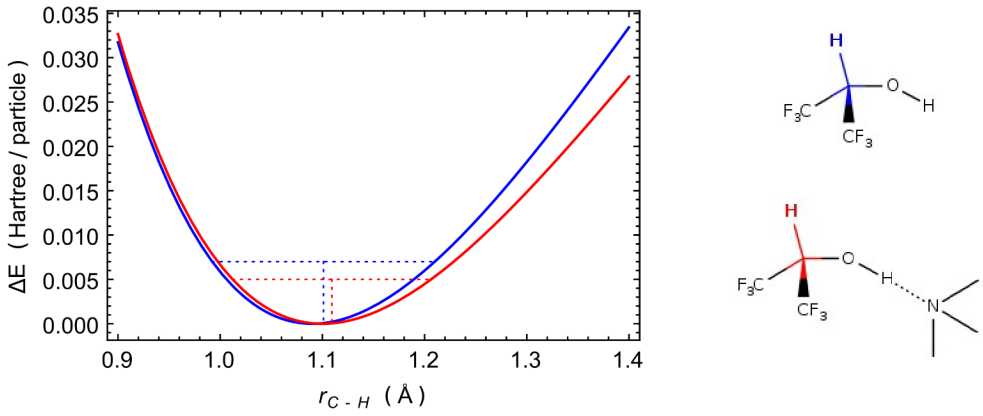
Finding the binding regions in disordered proteins are described in the chapter three. This chapter detailed the bioinformatics analysis of binding regions

found in the human intrinsically disordered proteins. The human proteome contains a significant number of intrinsically disordered proteins (IDPs). They show unusual structural features that enable them to participate in diverse cellular functions and play significant roles in cell signaling and reorganization processes. In addition, the actions of IDPs, their functional cooperativity, conformational alterations and folding often accompany binding to a target macromolecule. Applying bioinformatics approaches and with the aid of statistical methodologies, we investigated the statistical parameters of the binding regions (BRs) found in human disordered proteome. Firstly, we constructed a 471 leaf phylogenetic tree of human disordered proteome from multiple sequence alignment and analyzed the similarity and evolutionary distances between the IDPs. Distribution of the proteins with different degrees of disorderedness were also shown among the different clades of the disordered proteome. Further, the statistical models for the occurrence of BRs, their length distribution and percent occupancy in the parent proteins has been shown. The frequency of BRs followed a Poisson distribution pattern with increasing expectancy with the degree of increasing disorderedness. The length of the individual BRs also followed Poisson distribution with a mean of six residues, whereas, percentage of residues in BR showed a normal distribution pattern. We also explored the physicochemical properties such as the grand average of hydropathy (GRAVY) and the theoretical isoelectric points (pIs) of the BRs and compared them with that of the IDPs. The theoretical pIs of the BRs followed a bimodal distribution as in the parent proteins. However, the mean acidic/basic pIs were significantly lower/higher than that of the IDPs, respectively. We further showed that the amino acid composition of BRs was enriched in hydrophobic residues such as Ala, Val, Ile, Leu and Phe compared to the average sequence content of the IDPs. Conformational adaptibility of the BRs was also explored. Sequences in a BR showed conformational preference mostly towards flexible coil structure followed by helix, however, the ordered secondary structural conformation is significantly lower in BRs than the IDPs. Combining and comparing these statistical information of BRs with other methods may be useful for high-throughput functional annotation of proteins, drug target identification and drug discovery linking protein disorder.



When a small molecule enters into the binding site of a protein, what kind of microenvironment it experiences from the bulk solvent and how that environment affects the physical properties of the ligand molecule has been addressed in the chapter four. This chapter described the development of a fluorescent probe molecule and how that probe molecule can be useful to explore the hydrophobic binding regions in a target protein. Fluorescence emission and anisotropy are widely used to measure the binding parameters and kinetic behavior of reactions that cause a change in the rotational time of a fluorescent molecule. The fluorescence emission and anisotropy behavior of the newly synthesized novel naphthalene base fluorophore (methyl 3-[(6-{[2-(tert-butoxy)-2-oxoethyl] (4-methoxyphenyl)amino}naphthalen-2-yl)formamido]propanoate) in several solution conditions including its binding to human and bovine serum albumins, both in their native and denatured states has been discussed in this chapter. The fluorescence yield of the compound substantially increased inside hydrophobic grooves of the target proteins and around 30 nm decrease in Stokes' shift, compared to aqueous solution, was observed. The shift in fluorescence excitation peak position from the absorption peak of the molecule was ~8 nm in protein environment, which indicated the possible alteration of excited state geometry of the compound by the globular fold of target proteins. In addition, the steady state fluorescence anisotropy of the molecule was measured to evaluate several thermodynamic parameters. The thermodynamic analysis suggested that the binding was

energetically favorable. The measured ΔG° was about -30 kJ mol^{-1} and the derived dissociation constant was ~10^{-6} M. The molecular docking analysis further highlighted the nonspecific nature of association of the compound with the proteins indicating that the hydrophobic forces played a significant role in the binding processes. Under the denatured condition of the protein, the compound lost its binding efficacy and reduction in heightened fluorescence intensity was observed suggesting the release of the probe into the solvent. Thus, the molecule appears as a new fluorescence probe to report the nature of its binding site in terms of increased fluorescence quantum yield and decreased Stokes' shift. It can also report the changes in the binding site due to global change in protein structure such as unfolding/misfolding often linked to several human disorder. Further it could be useful to detect and study the drug binding site of specific protein of interest.



In the last chapter, a way to probe the strength of hydrogen bonding using an isotope labeling method has been discussed. Experimental measurement of contribution of hydrogen bonding to intermolecular and intramolecular interactions that provide specificity to biological complex formation is an important aspect of macromolecular chemistry and structural biology. However, there are a very few viable methods available to determine the energetic contribution of individual hydrogen bond to binding and catalysis in biological systems. Therefore, the methods that use secondary deuterium isotope effects analyzed by NMR or equilibrium or kinetic isotope effect measurements are attractive ways to gain

information on the hydrogen bonding properties of an alcohol system, particularly in biological environment. In this chapter, the anharmonic contribution to the C-H group was explored when the O-H group of 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) form intermolecular hydrogen bond with the amines by quantum mechanical calculations and by experimentally measuring the H/D effect by NMR. Within the framework of density functional theory, *ab initio* calculations were carried out for HFIP in its two different conformational states and their H-bonded complexes with tertiary amines to determine the ^{13}C chemical shielding, change in their vibrational equilibrium distances and the deuterium isotope effect on $^{13}C2$ (secondary carbon) of HFIP upon formation of complexes with tertiary amines. When C2-OH involved in hydrogen bond formation (O-H as hydrogen donor), it weakened the geminal C2-H bond; it was reflected in the NMR chemical shift, coupling constant and the equilibrium distances of the C-H bond. The first derivative of nuclear shielding at C2 in HFIP was -48.94 and -50.73 ppm $Å^{-1}$ for anti and gauche conformations, respectively. In the complex, the values were -50.28 and -50.76 ppm $Å^{-1}$, respectively. The C-H stretching frequency was lower than the free monomer indicating enhanced anharmonicity in the C-H bond in the complex form. In chloroform, HFIP formed complex with the amine; δC2 was 69.107 ppm for HFIP-tryethylamine and 68.766 ppm for HFIP-d2-tryethylamine and the difference in chemical shift, the ΔδC2 was 341 ppb. The enhanced anharmonicity in the hydrogen bonded complex resulted larger vibrational equilibrium distance in C-H/D bonds. An analysis with Morse potential function indicated that the enhanced anharmonicity encountered in the bond was the origin of larger isotope effect and the equilibrium distances. Change in vibrational equilibrium distance and the deuterium isotope effect, as observed in the complex, could be used as parameters in monitoring the strength of the hydrogen bond in small model system with promising application in bio-macromolecules.

ACKNOWLEDGEMENTS

"I can no other answer make but thanks,
And thanks; and ever thanks..."
William Shakespeare

Firstly I thank Dr. Maiti, who gave me academic guidance, freedom as well as all support I could wish for to finish this project. Defining the task is limiting one's potential. Over the years I got involved in various projects of various research groups. I worked on malaria, breast cancer, gastric ulcer, gene regulation, DNA cleavage, nanoparticles and organic chemistry. Such a wide exposure developed my perspective to look into a problem. I doubt if I could get such liberty to work anywhere else. Thanks to Prof. Abhijit Chakraborty and Prof. Samita Basu. My journey of the study of protein and small molecule interactions started from their labs at the Saha Institute of Nuclear Physics. Prof. Basu used to teach us spectroscopy every afternoon when I was a summer trainee there. The seed that was sown then has grown over the years and the roots reached further.

I thank my lab-mates and co-authors Swagata, Mritunjoy, Supriya and Sandip for working with me in one of my project described in chapter three. I thank my lab-mate and co-author Sudeshna. She was involved in a project which constitutes the last chapter of my thesis. Thanks to my lab-mates Anupam and Kaushik for many interesting discussions. Thanks to my co-author Nitin, a trainee from NIPER, Kolkata, who worked with me in the disordered protein related project. Thanks to my co-author Baisali, a masters' student from University of Calcutta, who worked with me during her summer internship in my project "Probing the micro-environment of binding site with small molecules." Thanks to Dr. Sumit Pramanik and Dr. Biswadip Banerji with whom I published several papers including the published work described in Chapter four of this thesis. It was always been a pleasure to work with Sumit Pramanik. I was always impressed by his capacity to work hard with no apparent stress.

Thanks to Dr. Vernon E. Anderson whose encouragement helped to initiate the study on hydrogen bond strength. I thank Prof. Samaresh Mitra for many interesting and stimulating discussions and for his suggestion to make the thesis concise. I hope it is not too long. I Thank Dr. Snehasikta Swarnakar, Dr. Uday Bandyopadhyay and Dr. Mrinal Kanti Ghosh for including me in their various drug development projects, which helped me a lot to improve my work on developing computational methods to study the intricacies of protein small molecule interactions.

The money came from the Department of Science and Technology of the Government of India. In spite of the long delays in fund release process and the interminable wait into pennilessness, I sincerely thank Department of Science and Technology for the financial support they provided through their INSPIRE fellowship programme. I am also grateful to the Council of Scientific and Industrial Research/Indian Institute of Chemical Biology for the provision of chemicals, instruments and the workspace.

Finally, I have to express my gratitude to my mother. She is my only living family. Except the cats, of course. She has always supported me during all these years. I couldn't have come this far without her support. I lost my father too early, but I know he would have been proud of me. He was an altruistic person. In the village where I grew up, everybody knew me because of him and I used to boast about that. One day, he told me that the people who were good to me because of him would turn their backs when he was not there anymore. So, he told me not to live in his shadows but to make my own mark. He didn't want me to be known as his son; rather, he wished to be known for being my father.

CHAPTER 1 | FINDING SPECIFICITY IN PROTEIN-SMALL MOLECULE INTERACTIONS: A STRUCTURE BASED COMPUTATIONAL APPROACH

Keywords: virtual screening, docking, randomness, skew normal distribution

INTRODUCTION

Importance of computational drug discovery. Finding the right molecule for the right target is a challenge in drug discovery because there are infinite number of potential drug like compounds for screening against a vast biological target. Chemical space is enormous. It has been estimated that there could be 10^{60} synthetically feasible organic molecules of molecular weight 500 or less (Barker et al., 2013; Rupakheti et al., 2015; Virshup et al., 2013). On the other hand, there are thousands of validated and tractable targets known to choose from (Hopkins and Groom, 2002; Kim et al., 2016; Overington et al., 2006). Therefore, even with the advancement of combinatorial chemistry and high throughput methods, it is a futile effort to screen the chemical space in search of the right molecule. It still needs our careful consideration for what to screen or to make. The chemical universe database GDB-17 lists 166.4 billion molecules of up to 17 atoms of C, N, O, S and halogens (Reymond, 2015; Ruddigkeit et al., 2012), which is four order of magnitude more than the number of known molecules in that size range (Kim et al., 2016; Virshup et al., 2013). However, a typical pharmaceutical compound collection does not even match these numbers. High throughput screenings allow about a million compounds to be tested, which is a very small proportion of the available drug like compounds (Kogej et al., 2013; Schamberger et al., 2011). Besides, large scale screening is expensive and not all the targets are suitable for high throughput screenings. These shortcomings in drug development led to the necessity of virtual screening, which refers to a range of *in silico* techniques used to search large compound databases to select a smaller number for biological testing.

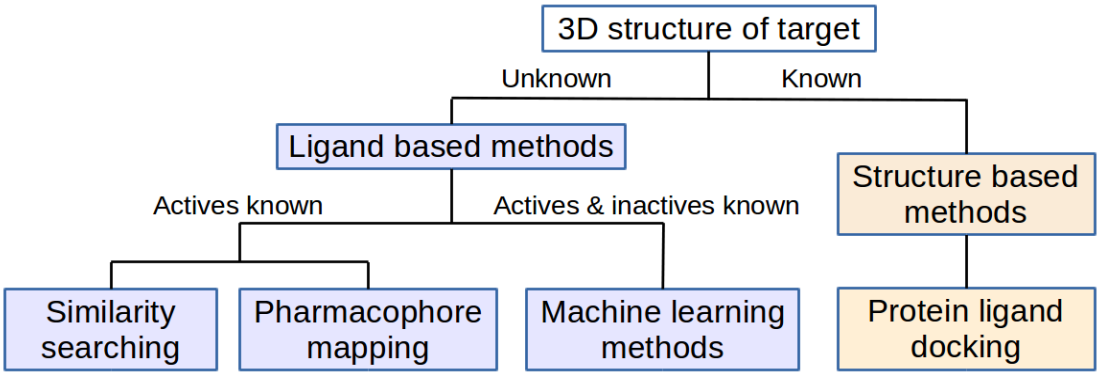


Figure 1-1: Methods in virtual screening. Different ligand based or structure based approaches are taken depending on the knowledge of target structure and/or active/inactive ligands.

Different approaches of computational drug discovery. The technique applied in virtual screening depends on the amount of information available about the particular disease target or the active ligands. Figure 1-1 summarizes the methods in virtual screening while figure 1-2 compares these methods on a predictivity-interpretability scale. When the 3D structure of the target is unknown, ligand based approach is taken. Similarity search and pharmacophore mapping are two such popular ligand based methods (Leach and Gillet, 2007a; Willett, 2006). The similar property principle states that structurally similar molecules tend to have similar properties, e.g., morphine, heroin and codaine are structurally similar and exerts similar pharmaceutical effect (Johnson et al., 1990). The similarity based method has been the most predictive method so far in the rational drug discovery (Brown and Lewis, 2006; Willett, 2006). However, structurally unrelated compounds are also known to produce similar effects such as methadone, which is only 20% similar to morphine but produces similar effect. Moreover, the existence of "activity cliffs", which suggests that relatively small structural changes cause relatively large potency changes, rendered this method least interpretable (Cruz-Monteagudo et al., 2014; Hu and Bajorath, 2012; Medina-Franco, 2013). Nevertheless, similarity search formed the basis of medicinal chemistry efforts and of all ligand based virtual screening methods (Willett, 2006). Another ligand based approach is pharmacophore mapping. A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal interactions with a specific target protein and to modulate its biological response (Leach et al., 2010).

This method is relatively more interpretable than the similarity search as it provides information about the functional groups involved and the probable types of interaction (Brown and Lewis, 2006). This method can also be applied in a structure based drug design where the 3D structure of the target is known. Then, there is the structure activity relationship (SAR) modeling which make use of the knowledge of inactive ligands as well as the actives in the ligand based drug design (Leach and Gillet, 2007b). SAR and the quantitative SAR (QSAR) heavily relies on machine learning and is a more rational approach in drug development. And finally, the most interpretable but least predictive approach is the structure based drug design: protein small molecule docking.

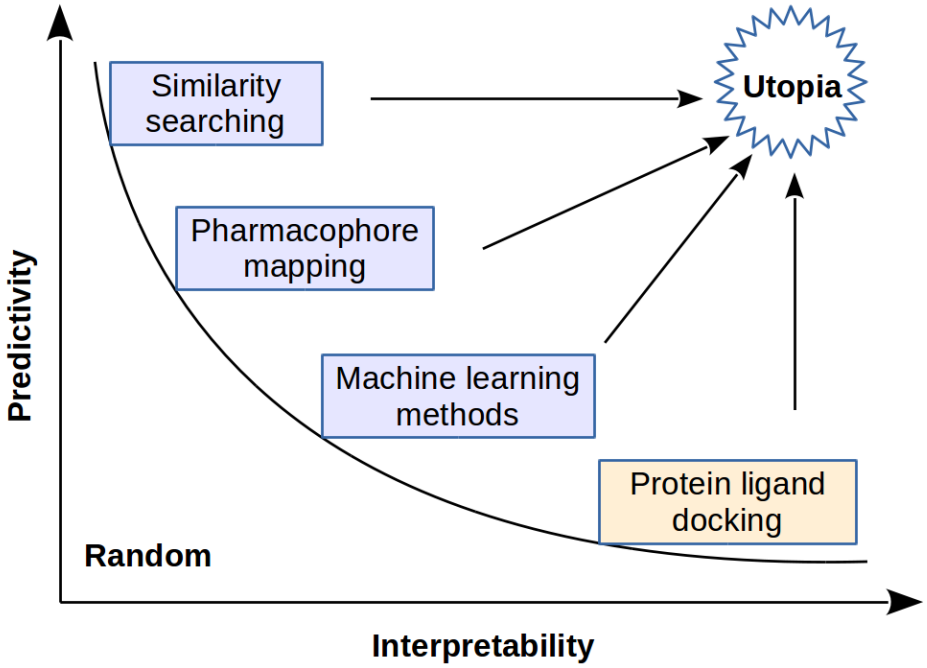


Figure 1-2: Predictive modeling. Different approaches in computational drug discovery are compared on a predictivity and interpretability scale.

Docking in retrospect. The first protein crystal structure was solved in 1958, however, its implication in drug discovery was not realized until later (Kendrew et al., 1958). The ability to produce large amounts of proteins from cloned genes in the 1970s dramatically expanded the range of protein structures that could be determined (Jackson et al., 1972). In the early 80s, protein structure determination became an accepted tool to facilitate drug discovery. In the year

1982, the first docking program, DOCK was introduced (Kuntz et al., 1982). The program was relatively simple and relied on the lock and key model of enzyme action and used shape complementarity to position the rigid ligand withing the active site. Over the years, many docking programs with complex algorithms have been developed, which explores the conformational space of ligand at run time and also allows partial flexibility of the receptor side chains in the binding site (Friesner et al., 2004; McGann, 2011; Morris et al., 2009; Rarey et al., 1996; Taylor et al., 2002; Trott and Olson, 2010; Venkatachalam et al., 2003; Verdonk et al., 2003). However, all the docking algorithms are based on two key components: search algorithm and scoring function. The search algorithm is used to generate conformation, position and orientation of the ligand within the active site and to orient the amino acid side chains in the binding cavity as well. Whereas, the scoring function is used to estimate binding affinity in order to identify the most likely pose for an individual ligand and to assign a priority order to a set of diverse ligands docked to the same protein. Current algorithms of docking include fragment based methods (FlexX), Monte Carlo/simulated annealing (AutoDock, Affinity, LigandFit), genetic algorithms (GOLD, AutoDock) and systematic search (FRED, Glide) (Friesner et al., 2004; Jones et al., 1997; Kramer et al., 1999; McGann, 2011; Morris et al., 2009; Venkatachalam et al., 2003). In the fragment based method, the ligand is broken down into fragments and then reconstructed step wise within the active site (Kramer et al., 1999). Genetic algorithm, on the other hand, mimics the process of evolution (Clark and Westhead, 1996). In this method the search space, which is defined encompassing the binding site or the whole protein, is populated with the ligand carrying various mutations (active torsions, translation and rotation in three dimensions) and Only the fittest (the best affinity conformer) survives into the next generation; the cycle goes on until the specified number of iteration is reached. Performance of these algorithms varies from target to target, and scoring function to scoring function. Docking is the most interpretable method in drug design and reasonably good at finding the correct pose for a given protein ligand complex (native docking). However, it is less good at ranking different ligands against the same protein (virtual screening) and even less good at finding the right target for a given ligand (inverse docking).

Utopia of computer aided drug discovery. In computational drug discovery, a method which is highly predictive as well as very interpretable would be the most ideal method. If such a method existed, drug development would have been a much simpler affair. However, none of the available methods fulfill that criteria. Similarity search is predictive but not so interpretable, whereas, the most interpretable structure based methods produce a lot of false positive. New techniques are being developed combining different approaches to improve the results and the performance varies on different datasets. Our aim was to improve the predictability of docking methods, thus, reducing the number of false positive hits in virtual screening. We incorporated some basic ideas, which are fundamental in our understanding of probability theory and statistical mechanics to develop computational filters in order to remove undesirable compounds from further consideration.

THEORY AND THE HYPOTHESIS

Docking and stochastic models. Due to the high degrees of freedom, systematic exhaustive search of the energy landscape to find the most favorable binding conformation is computationally costly and the quality of the result depends on the granularity of the sampling. Still, some docking programs such as Glide, and FRED use the systematic search approach (Friesner et al., 2004; Jones et al., 1997)to dock rigid ligands or precomputed rotamers of a flexible ligand, thus, restricting the degrees of freedom only to the rotation and translation along the three Cartesian axes. However, most of the docking programs uses stochastic (random) search models. In probability theory, a stochastic process is the probabilistic counterpart to a deterministic process. Instead of describing a process which can only evolve in one way, in a stochastic process there is some indeterminacy: even if the initial condition (or starting point) is known, there are several (often infinitely many) directions in which the process may evolve. Monte Carlo simulated annealing and genetic algorithms are the examples of such stochastic search models used by the most cited docking program AutoDock

(Morris et al., 2009; Sousa et al., 2006) to explore the conformational space of a ligand in combination with various scoring functions. Another popular docking program, GOLD (genetic optimization for ligand docking) is also based on the genetic search algorithm (Jones et al., 1997). Even the search algorithms of the fragment based docking programs DOCK and FlexX, which applies systematic search to incrementally build the ligand from fragments inside the binding site (Leach and Kuntz, 1992; Rarey et al., 1996), are a class of non-diffusion stochastic models. Such models have a hybrid state with two components. The first part of the state visits a finite set of modes in a stochastic manner. When the mode is fixed, the continuous state evolves in a deterministic manner (Shanbhag and Rao, 2003). Deterministic process is reproducible, whereas stochastic algorithms include random factors that do not allow full reproducibility. Therefore, repeated docking with the same ligand may produce different result each time the algorithm runs. Many programs such as AutoDock has the provision to run genetic algorithm multiple times to produce an ensemble of solutions allowing statistical mechanics analysis on the ensemble of solutions possible. In AutoDock, conformers are clustered according to their spatial distribution (in terms of root mean square deviation, RMSD) and energy (Morris et al., 2009); the lowest energy and/or highest frequency clusters are generally accepted as the solution. However, there are scope for further analysis such as how to extract the information about the specificity of interaction hidden in these clusters.

Underlying rhythm of randomness. Another important theory regarding our study of specificity is the distribution of certain random variables and how that distribution is influenced by different factors. One of the most common model of randomness is the normal distribution. For example, rabbits may have all possible wights, but, if we plan to record the wight of every single rabbit, we shall notice that most of the rabbits are close to the average. There are some rabbits, but not many, that are much larger or much smaller than the average. The further from the average we go, fewer animals are there. This particular bell shaped distribution, that centers around the mean, is called a normal distribution. However, many distributions that occur in practical situations are skewed, not symmetric. For

example, age at death is negatively skewed (Robertson and Allison, 2012) in developed countries because of the better health care systems available to the people in those countries. Certain quantities in physics are distributed normally, as was first demonstrated by James Clerk Maxwell. Examples of such quantities include velocities of the particles in any system in thermodynamic equilibrium and the position of a particle that experiences diffusion (Maxwell, 1860a, 1860b). However, the Maxwell-Boltzmann distribution applies only to the classical ideal gas (non-interacting particles) and it may be influenced by different factors such as the temperature of the system and the mass of the particle. In real gases, there are various effects such as van der Waals interactions and quantum exchange interactions, that make their speed distribution sometimes very different from the Maxwell-Boltzmann form. Similarly, study of the distribution of system states (conformers) generated by the repeated independent molecular docking iterations might provide useful information about the influences of specific interactions on ligand receptor docking.

Specificity in docking. Based on the theories discussed above, we hypothesized that specificity in the protein small molecule interactions can be derived from molecular docking experiments by implementing concepts abstracted from statistical mechanics, namely, the populations. The lack of reproducibility in molecular docking prediction allows generation of an ensemble of conformers (population) by repeated independent docking calculations on a single protein ligand complex. The conformers in that ensemble generated under randomized trials may adopt all the possible energies and spatial orientations/position within the search space (a statistical ensemble). And our hypothesis is that the distribution of the system states (conformers) in that statistical ensemble is normal unless it is biased by the specific interactions between the interacting partners and that the deviation from the normality in the distribution of system states can be correlated to the degree of specificity in the interaction. We called it the specificity hypothesis. Moreover, the parameters of the distribution of system states such as the mean, deviation and skewness can be used to filter false positive hits in the virtual screening work flow.

Origin of specificity. We discussed above that the distribution of the system states obtained by repeated independent molecular docking experiments follow a normal distribution unless specific interactions between the interacting partners bias the distribution toward a particular binding mode. Based on this idea we can say that the origin of specificity points to the pharmacophoric features of the binding site and complementarity in ligand. A pharmacohore is an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response (Wermuth et al., 2009). To better understand this, two ideal conditions may be considers. Firstly, we can imagine a situation where the ligand and its binding site is devoid of any functional groups that can take part in some specific type of interactions such as the hydrogen bonding. We can further assume that the ligand and the binding site does not share any shape complementarity. Under such a situation, only the non specific interactions, such as van der Waal's and hydrophobic interactions are possible between the ligand and the protein, which allows the ligand to take any orientation and position within the binding site of the protein without much fluctuations in binding energy. Now, if we distribute the system states generated by docking iterations for this ligand receptor pair, we should observe a normal distribution as shown in the figure 1-3.

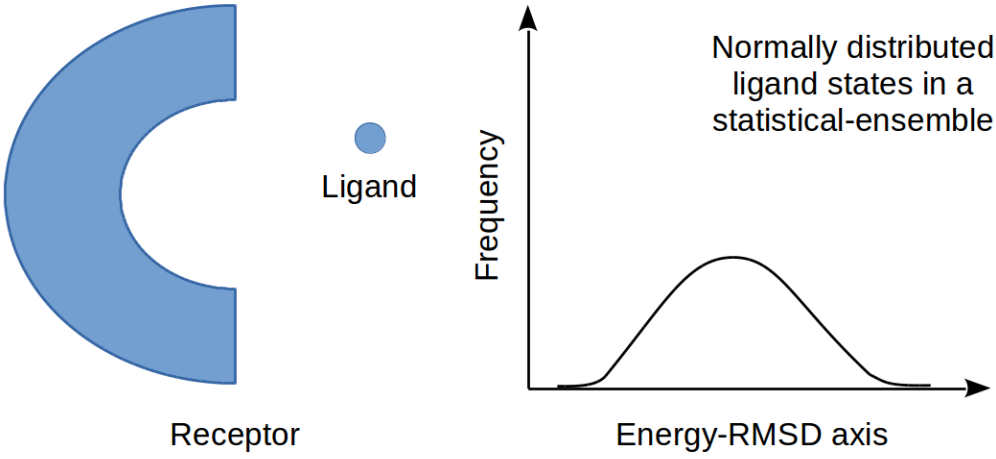


Figure 1-3: Non-specific interactions. The receptor and the ligand does not show any elements of complementarity. Repeated docking under randomized condition generates system states which are normally distributed.

In the second hypothetical setup, some elements of complementarity, such as a hydrogen bond donor in the ligand and an acceptor in the receptor was introduced as shown in the figure 1-4. In such a setup, the hydrogen bonded complex formation between the ligand and the receptor would be preferred as it is thermodynamically more favorable. Therefore, the ensemble of docked conformations would be more populated with the hydrogen bonded conformers of the ligand receptor complex. Thus, the specific interactions can bias the distribution of system states toward a specific low energy conformation resulting in a positive skewness of the distribution as shown in figure 1-4. This skewness suggests that the degree of bias is quantifiable and can be used as a filter in virtual screening to improve the reliability of docking simulation many folds. Moreover, as the skewness directly correlates with the specificity of interaction, it can be called a specificity parameter, α. Another important parameter that can be obtained from this skewed distribution is the spread of the distribution, σ (Figure 1-4). Lower the value of the spread indicates higher the probability of finding a specific low energy conformer in the ensemble. Thus, the spread is inversely proportional with the specificity. The spread also accounts for reproducibility of the docking experiment and, therefore, may be called a reproducibility parameter.

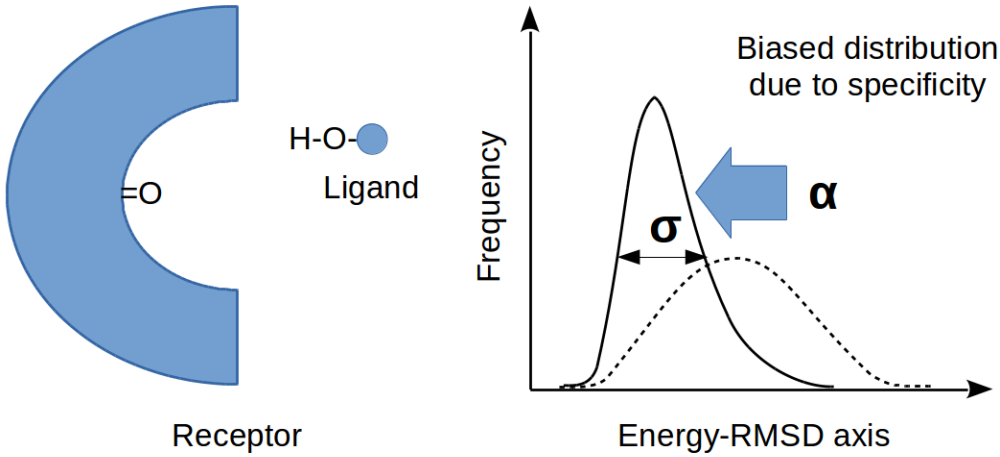


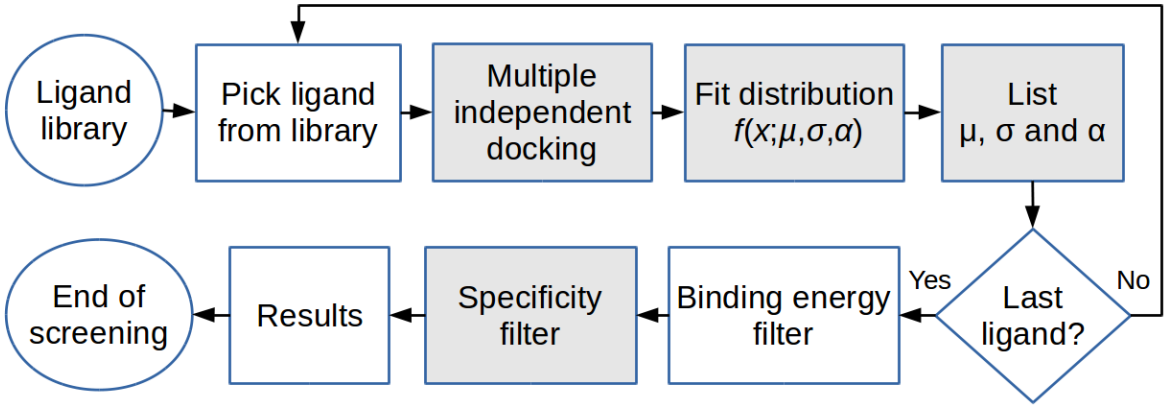
Figure 1-4: Specificity and reproducibility. The receptor and the ligand shows an element of complementarity (a hydrogen bond acceptor in the receptor and donor in the ligand). Specific interactions such as hydrogen bonding can significantly bias the distribution toward a lower energy state. α is a measurable parameter of specificity. σ is a reproducibility parameter.

MATERIALS AND METHODS

In oder to test the hypothesis, a protocol of virtual screening was developed. The work flow was based on the conventional docking programs, however, it incorporated a specificity filter. The scheme of the protocol is shown in the figure 1-5. According to this protocol a ligand is picked up from a library of small molecules and then docked onto the target protein. The docking runs in a loop until a specified number of iterations is reached. Each docking calculation in this loop is independent of the previous one. After a certain number of docked conformations are generated, the conformers are clustered according to their binding energy, spatial orientation and position. Distribution of the frequencies of these clusters is then analyzed by fitting a skew normal distribution with location parameter μ, spread σ and skewness parameter α (Figure 1-4 and 1-5). The values of the μ, σ and α are listed and the whole process is repeated for the next ligand in the library. Finally, the ligands are scored according to the values of μ, σ and α. The thermodynamic parameter, μ, is used in traditional binding energy filters to score the ligands. Whereas, σ and α forms the specificity filter as they are directly or inversely related to the specificity of the interactions, respectively. Higher value of α, which is generally associated with the lower value of σ indicates very specific interactions.

AutoDock 4.2 of The Scripps Research Institute was used as the base program in this method (Morris et al., 2009). The structures of the proteins were obtained from the Protein Data Bank (Berman et al., 2000) and the ligand molecules were either obtained from PubChem (Kim et al., 2016) or drawn on Avogadro (Hanwell et al., 2012) followed by geometry optimization in molecular mechanics force filed using the steepest descent algorithm. Proteins and the ligands were preprocessed in AutoDockTools (Morris et al., 2009) for docking, where the non-polar hydrogens were merged, Gaisteiger charges were added and the atom types were defined. Any water, duplicate atoms or undesirable chemical entities were also removed from the protein structure files. Rotatable bonds in the ligands were defined for the flexible docking. Search spaces were defined by

encompassing the whole protein or the part of it where the binding pocket was. Lamarckian genetic algorithm with Solis & Wets' local search methods was used as the search algorithm (Morris et al., 1998). Genetic algorithm runs were set to 100 or 1000 with 25 millions energy evaluations each time. The docked conformation were clustered with energy cut offs of 0.5 kcal per mole and 2 Å spatial deviation (RMSD). Frequency of these clusters were then distributed over the energy axis and the skew normal distribution with parameters μ, σ and α was fitted to it in Wolfram Mathematica. Ligands were ranked according to their specificity and the binding energy for the comparison of this method with the traditional binding energy based scoring functions.



$$f(x;\mu,\sigma,\alpha) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} Erfc\left[\frac{-\alpha(x-\mu)}{\sigma\sqrt{2\pi}}\right]$$

Figure 1-5: Modified work flow of virtual screening. Specificity filter added to reduce the number of false positive hits. The additional steps required are highlighted in gray. The equation describes a skew normal distribution with mean μ, standard deviation σ and skewness parameter α. *f*(x) is the expected value of the random variable x. Erfc is the complementary error function.

RESULTS AND DISCUSSION

This protocol described here has already been implemented to study a wide range of protein ligand complexes in our collaborative drug development projects and the results were published in various international peer reviewed journals. The

proteins we studied include: a pro-inflammatory cytokine, macrophage migration inhibitory factor (MIF) and its two orthologues from *Plasmodium falciparum* and *Plasmodium yoelii*, which are potential targets for anti-malarial drug development (Alam et al., 2011, 2012); kinase domain of epidermal growth factor receptor (EGFR), which is involved in breast cancer (Bhowmik et al., 2013); matrixmetalloprotease-9, which is involved in stress related gastric ulcer and cancer metastatis (Rudra et al., 2012); and the transport proteins of serum, the serum albumin and its bovine orthologue (Banerji et al., 2013a; Pal et al., 2015). The small molecules include organic synthetic compounds (Alam et al., 2011; Bhowmik et al., 2013; Pal et al., 2015), small synthetic peptides (Banerji et al., 2013a), compounds extracted from natural sources (Alam et al., 2012) and an indigenous hormone (Rudra et al., 2012) as shown in the figure 1-6. Here, some of the published results produced by myself has been highlighted solely from the perspective of protein small molecule interactions.

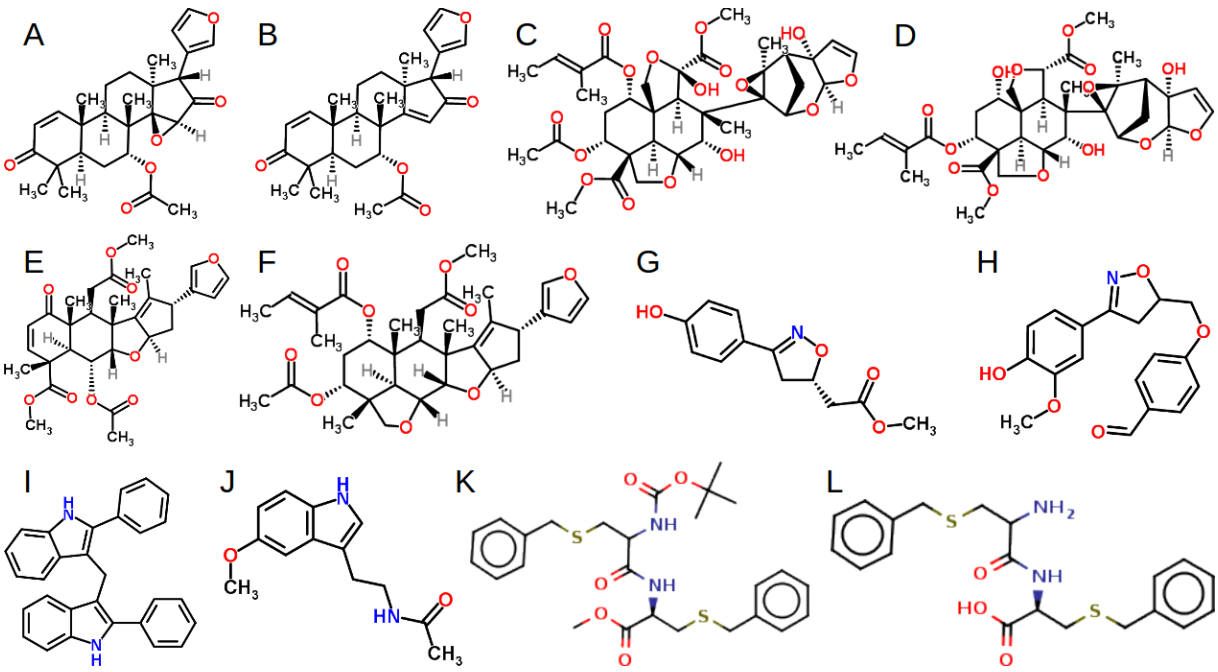


Figure 1-6: Small molecules used in this study. (A) Epoxyazadiradione. (B) Azadiradione. (C) Azadirachtin A. (D) Azadirachtin B. (E) Nimbin. (F) Salannin. (G) 4,5-dihydro-3-(4-hydroxyphenyl)-5-isoxazoleacetic acid methyl ester (ISO-1). (H) 4-((3-(4-Hydroxy-3-methoxyphenyl)-4,5-dihydroisoxazol-5-yl) methoxy) benzaldehyde (CP4b). (I) 2,2′-diphenyl-3,3′-diindolylmethane (DPDIM). (J) Melatonin. (K) Protected L-Cys-L-Cys/L-Cys-D-Cys (1A/1B). (L) Deprotected L-Cys-L-Cys/L-Cys-D-Cys (1C/1D). Compounds A—F are natural products obtained from the plan *Azadirachta indica*.

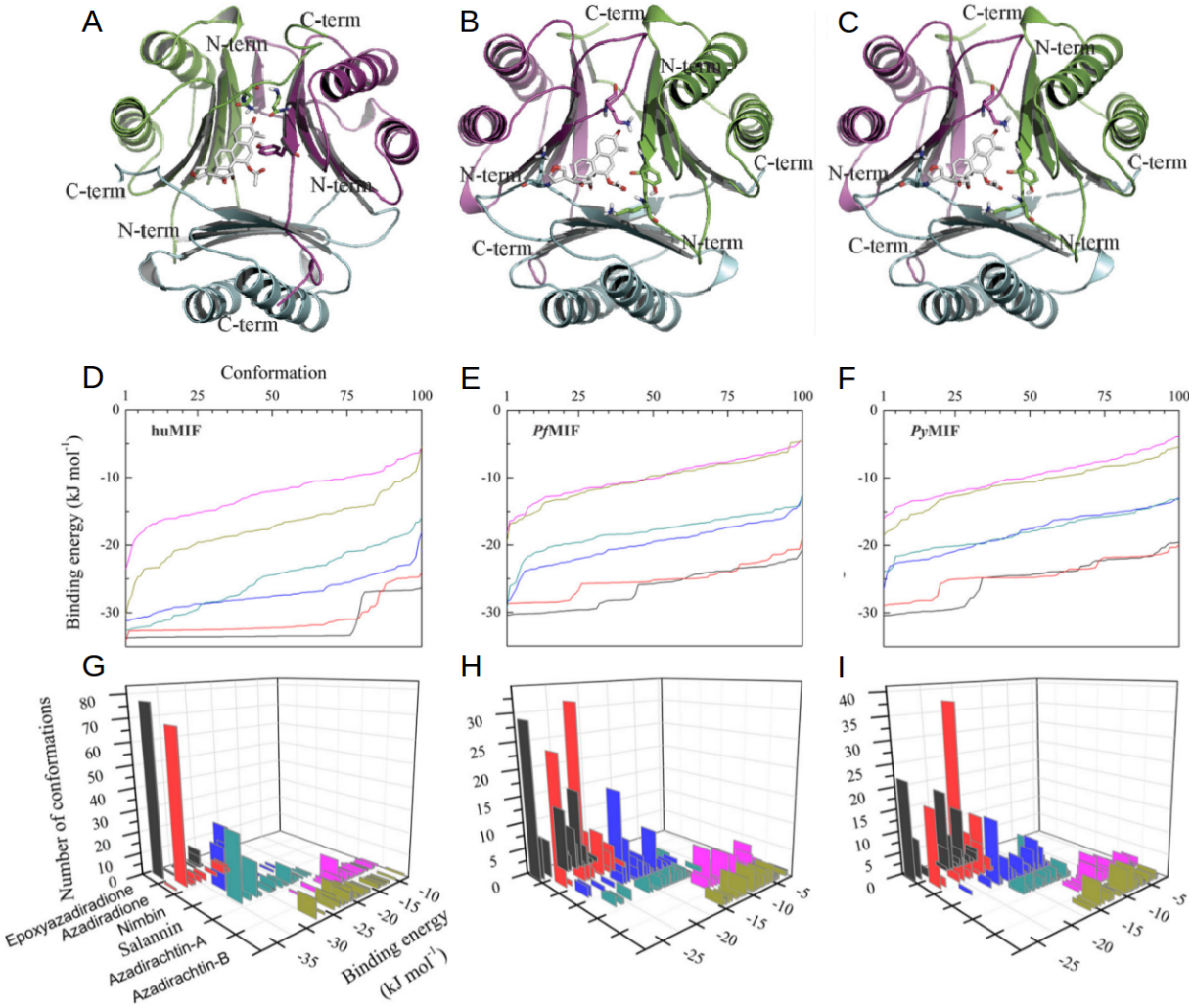


Figure 1-7: Interactions of limonoids with MIFs. (A) Human MIF complexed with epoxyazadiradione. (B) *Plasmodium falciparum* MIF bound to epoxyazadiradione. (C) *Plasmodium yoelii* MIF in complex with epoxyazadiradione. Proteins are shown in ribbon diagrams and the ligand in stick. Each monomer is indicated by a different color: chain A is green, chain B is cyan, and chain C is magenta and the ligand is white. Interacting side chains of amino acid residues are shown in sticks. (D-I) Binding energy analysis of limonoids. Color key: black, epoxyazadiradione; red, azadiradione; blue, nimbin; dark cyan, salannin; purple, azadirachtin A; dark yellow, azadirachtin B. (D-F) The energy spectrum distribution of system states (docked conformations of lmonoids) as determined by molecular docking for the targets: human, *Plasmodium falciparum, Plasmodium yoelii* MIFs, respectively. (G-I) The corresponding cluster distribution over the energy axis.

Some specific interactions. We first implemented this method to screen six compounds (limonoids) extracted and purified from the plant *Azadirachtca indica* (commonly known as Neem) for their anti-malarial activity (Alam et al., 2012). Their interactions with the macrophage migration inhibitory factor, a pro-inflammatory

cytokine and an anti-malarial target, was studied computationally (Figure 1-7). It was observed that all the six compounds produced very low energy binding modes with the three orthologous target proteins. However, when the distribution of the clusters were analyzed, it appeared that the binding of only two compounds, azadiradione and epoxyazadiradione, were highly specific. The two compounds are very similar differing only in one epoxy functional group. Both of the compounds produced narrow and highly positively skewed clusters, whereas, the distribution of the rest of the molecules were bell shaped or highly spreaded. Later, it was experimentally shown that only one of these two hits exerted inhibitory effect on the enzyme. Therefore, in this particular example we can see that the implementation of the specificity filter improved the success rate docking from one in six to one in two.

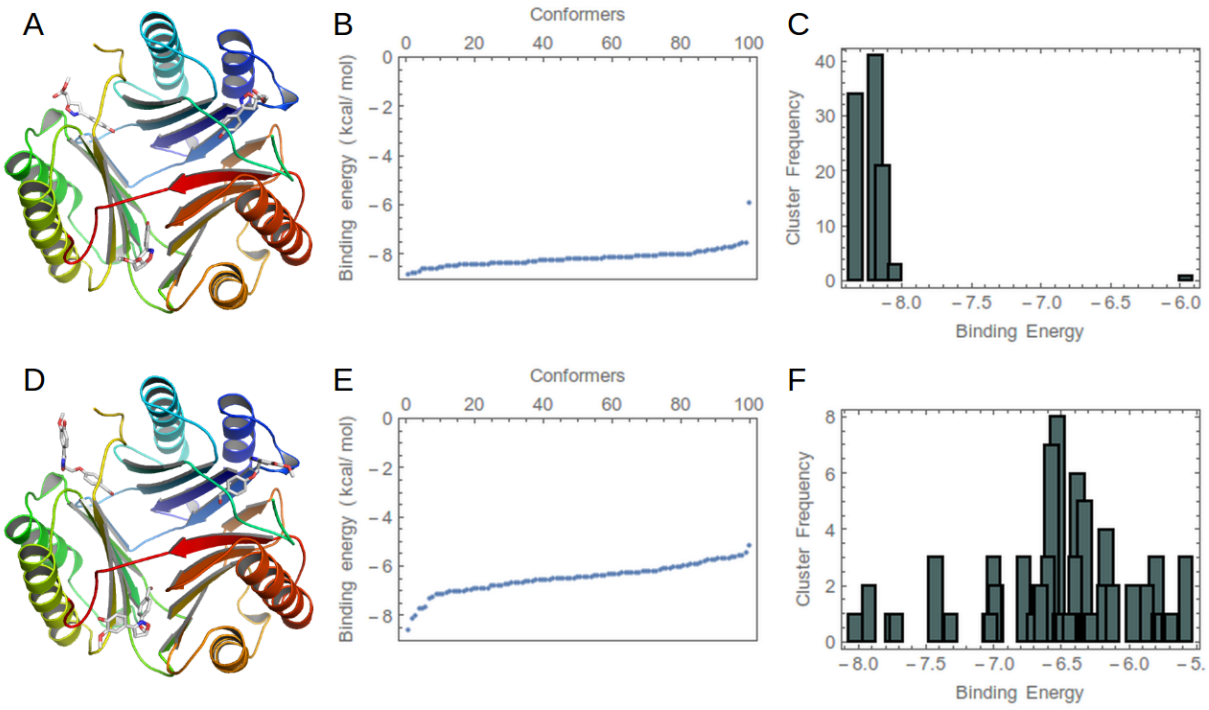


Figure 1-8: Interactions of ISO-1 and CP4b with human MIF. (A) Human MIF complexed with ISO-1. (B) The energy spectrum distribution of system states of ISO-1 as determined by molecular docking. (C) Cluster distribution over the energy axis. (D) Human MIF bound to CP4b. (E) The energy spectrum distribution of system states of CP4b as determined by molecular docking. (F) Cluster distribution over the energy axis. Proteins are shown in ribbon diagrams and the ligand in stick. Each monomer is indicated by a different color: chain A is blue, chain B is green, and chain C is red and the ligand is white.

The compound 4,5-dihydro-3-(4-hydroxyphenyl)-5-isoxazoleacetic acid methyl ester (ISO-1) is a known inhibitor of human MIF (Lubetsky et al., 2002). However, this experimental anti-inflammatory drug cannot inhibit the MIFs of *Plasmodium falciparum* or *Plasmodium yoelii*, which are potent anti-malarial targets and, thus, cannot be used as anti-malarial agent. With the aim to develop an inhibitor of plasmodial MIFs, 4-((3-(4-Hydroxy-3-methoxyphenyl)-4,5-dihydro isoxazol-5-yl) methoxy) benzaldehyde (CP4b) was designed (Alam et al., 2011). Although CP4b showed moderate inhibition on human MIF, it failed to inhibit the plasmodial MIFs. In docking analysis as well, CP4b showed binding energies for MIF comparable to that of ISO-1, however, the distribution of clusters was found to be bell shaped with a very high spread (Figure 1-8). Unlike CP4b, clusters distribution of ISO-1 was narrow and highly positively skewed, which suggested ISO-1 was a much better ligand than the CP4b. CP4b failed to pass the specificity filter.

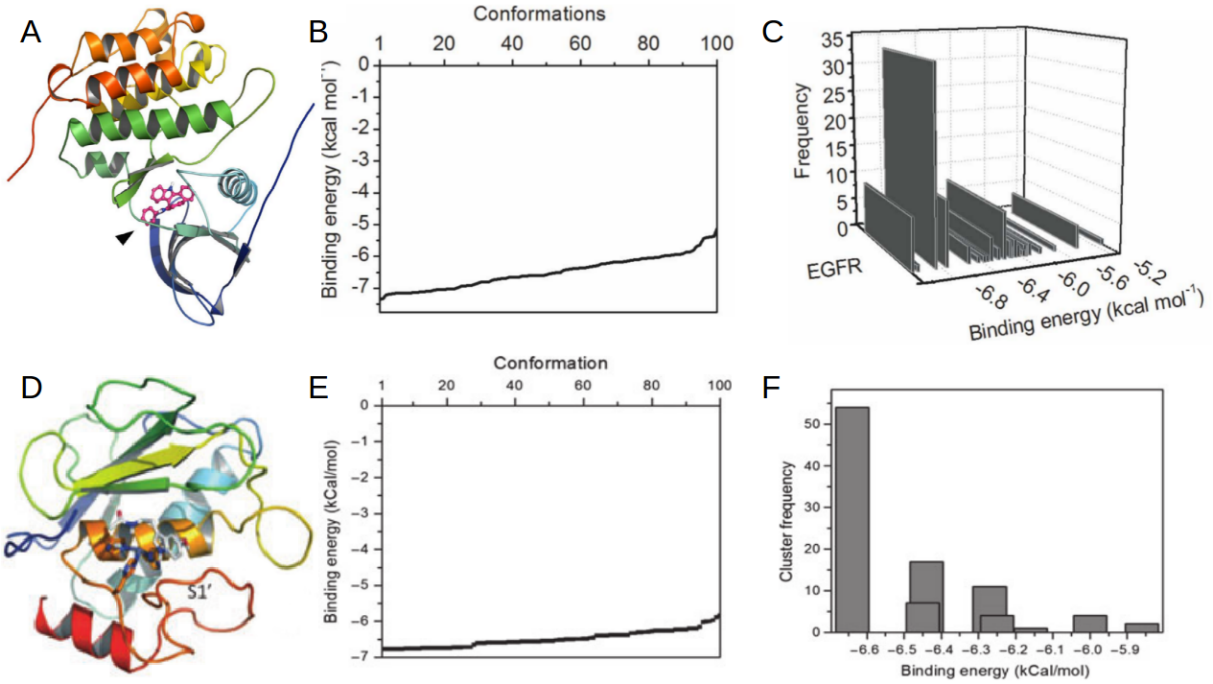


Figure 1-9: Binding of DPDIM to EGFR and melatonin to MMP-9. (A) EGFR complexed with DPDIM. (B) The energy spectrum distribution of system states as determined by molecular docking. (C) Cluster distribution over the energy axis. (D) MMP-9 bound to melatonin. (E) The energy spectrum distribution of system states as determined by molecular docking. (F) Cluster distribution over the energy axis. Proteins are shown in ribbon diagrams and the ligand in stick. N to C -terminal of the proteins are colored in rainbow and the ligand in white.

We further implemented this method to find 2,2′-diphenyl-3,3′-diindolylmethane (DPDIM) among several indole derivatives, which showed specific interaction with epidermal growth factor receptor (EGFR) kinase, both computationally (Figures 1-9A to 1-9C) and experimentally (Bhowmik et al., 2013). Indole based compounds are well known for inhibitors of matrixmetalloprotease-9 (MMP-9), which is involved in stress induced gastric ulcer and metastatic cancer. We found that melatonin, a hormone of pituitary gland and an indole derivative, can inhibit MMP-9 at low micromolar concentrations (Rudra et al., 2012) and the computational analysis also suggested specificity in the interaction (Figures 1-9D to 1-9F).

Non-specific interaction with serum albumins. Interaction of drugs with serum albumin is another major aspect of drug development. Serum albumins increase the solubility of hydrophobic ligands in plasma, modulates their delivery to cells, act as the reservoir of drugs and affects the bioavailability. Serum albumins have at least seven hydrophobic grooves on its surface, the precise architecture of which is known from several crystallographic and NMR spectroscopic studies (Curry et al., 1998). Serum albumins provide a unique environment for non-specific interactions and act as a universal receptor for many drug molecules.

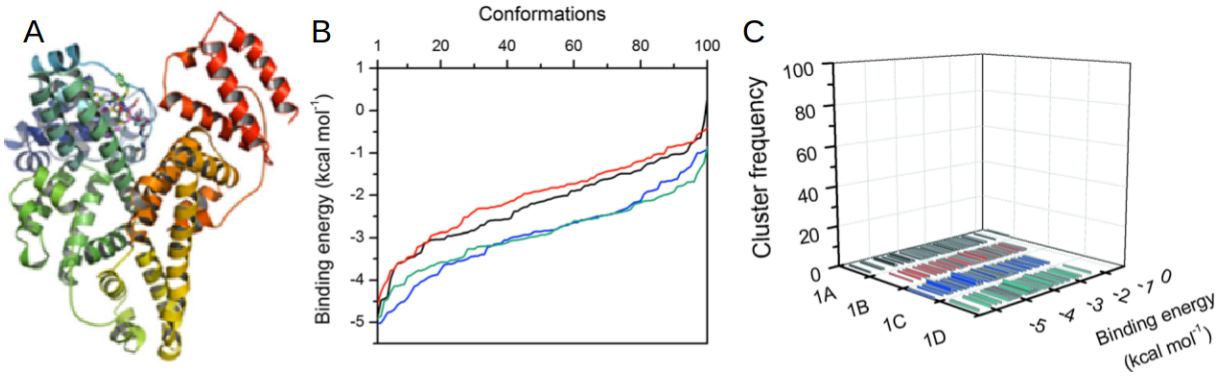


Figure 1-10: Binding of cystine-based dipeptides with serum albumin. (A) Human serum albumin complexed with cystine-based dipeptides, 1A-D. (B) The energy spectrum distribution of system states as determined by molecular docking. (C) Cluster distribution over the energy axis. Human serum albumin is shown in ribbon diagram and the ligands in stick. N to C -terminal of the protein is colored in rainbow and the four ligand (1A, 1B, 1C and 1D) are colored in green, magenta, light blue and cyan, respectively.

We have studied the interactions of newly synthesized peptides and small molecules with the serum albumins using experimental and computational methods (Banerji et al., 2013a; Pal et al., 2015). Computational analysis using the aforementioned approach produced cluster distribution suggesting non-specific nature of the interaction (Figure 1-10), which is in line with the experimentally established facts. The reproduction of such negative results validate, although indirectly, the approach to study specificity in the protein small molecule interactions by molecular docking.

CONCLUSIONS

We developed a method to find the specificity in the protein small molecule interactions. This approach was applied in smaller datasets and it was proved to be useful in finding the hits and reducing the false positives in virtual screening. However, this method needs to be standardized for bigger datasets and benchmarked against the library of decoys in order to implement in rational drug discovery pipeline.

CHAPTER 2 | FINDING THE BINDING SITES IN STRUCTURED PROTEINS: PATTERN BASED RECOGNITION OF ALLOSTERY AND DRUGGABILITY

Keywords: virtual screening, docking, ensemble analysis, clustering, density distribution

INTRODUCTION

Allostery. Allosteric modulators are ligands of proteins that bind in a binding site other than the site of the natural substrate. The site to which the ligand binds is termed the allosteric site. Allosteric regulations are a natural example of control loops, such as feedback from downstream products or feedforward from upstream substrates in a catalytic cascade. Allosteric regulation is also particularly important in the cell's ability to adjust enzyme activity by providing short-term, readily reversible regulation of metabolite flow in response to specific physiologic signals without altering gene expression. For example, lipogenesis (biosynthesis of fatty acids) is upregulated at the acetyl-CoA carboxylase step by the allosteric modifier citrate (Berg et al., 2002a), which is a key component of citric acid cycle and increases in concentration in the well-fed state indicating plentiful supply of acetyl-CoA. Citrate converts the enzyme acetyl-CoA carboxylase from an inactive dimer to an active polymeric form, having a molecular mass of several million. A classic example of allosteric modulation is the binding of oxygen to hemoglobin, where oxygen is effectively both the substrate and the effector (Berg et al., 2002b). The binding of oxygen to one subunit of hemoglobin induces a conformational change in the protein, which enhances oxygen affinity for the remaining binding sites facilitating subsequent binding with oxygen (positive cooperative binding). On the other hand, a metabolite of anaerobic glycolysis called 2,3- DPG (diphosphoglycerate) allosterically reduces the affinity of hemoglobin towards oxygen. Aspartate transcarbamoylase (ATCase) is another model allosteric enzyme (Berg et al., 2002b). ATCase, the catalyst for the first reaction unique to

pyrimidine biosynthesis (nucleotide metabolism), is feedback-inhibited by cytidine triphosphate (CTP). ATCase consists of multiple catalytic and regulatory subunits and each regulatory subunit contains at least two CTP (regulatory) sites. Figure 2-1 shows two enzymes bound to allosteric modulators. Fructose 1,6-bisphosphate, which a rate limiting enzyme in gluconeogenesis, is allosterically inhibited by adenosine monophosphate (Figure 2-1A). Adenosine phosphates are among the major allosteric effectors found in the body alog with nicotinamide adenine dinucleotides, acetyl- and acyl- CoA.

Moreover, the allosteric modulators are interesting drugs. Allosteric modulators can enable modulation of targets that are difficult to target orthosterically. There are a number of advantages in using allosteric modulators as preferred therapeutic agents over classic orthosteric ligands. For example, allosteric binding sites have not faced the same evolutionary pressure as orthosteric sites to accommodate an endogenous ligand, so are more diverse. Therefore, greater selectivity may be obtained by targeting allosteric sites (Christopoulos, 2002). This is particularly useful for targets such as GPCRs (Burford et al., 2011; Ivetac and Andrew McCammon, 2010) and kinases (De Smet et al., 2014; Hantschel et al., 2012; Zhang et al., 2010)where selective orthosteric therapy has been difficult due to the sequence conservation of the orthosteric site across receptor subtypes. In this respect, allosteric modulators are considered as emerging strategy in rational drug development. Recent advances in targeting protein kinases include the discovery of type III inhibitors that bind a site proximal to the ATP binding pocket as well as the truly allosteric type IV inhibitors that target protein kinases distal to the substrate binding pocket (Foda and Seeliger, 2014; Lamba and Ghosh, 2012; Liu and Gray, 2006). These new classes of inhibitors are often selective and usually display high affinities. Discovery and development of non-nucleoside reverse-transcriptase inhibitors (NNRTIs) are another example of allosteric drug development (Schauer et al., 2014; Seckler et al., 2011). NNRTIs are antiretroviral drugs used in the treatment of human immunodeficiency virus infection. NNRTIs are generally allosteric inhibitors of reverse transcriptase, an enzyme that controls the replication of the genetic material of HIV and is one of

the most popular targets in the field of antiretroviral drug development. Figure 2-1B shows the interaction of the allosteric NNRTI drug nevirapine with the HIV-1 reverse transcriptase (Das et al., 2014). Allosteric inhibitors act by binding noncompetitively or uncompetitively to the target protein, however, unlike the agonists or antagonists, allosteric modulators can also amplify signals (Schwartz and Holst, 2007) rather than activate/inhibit them.

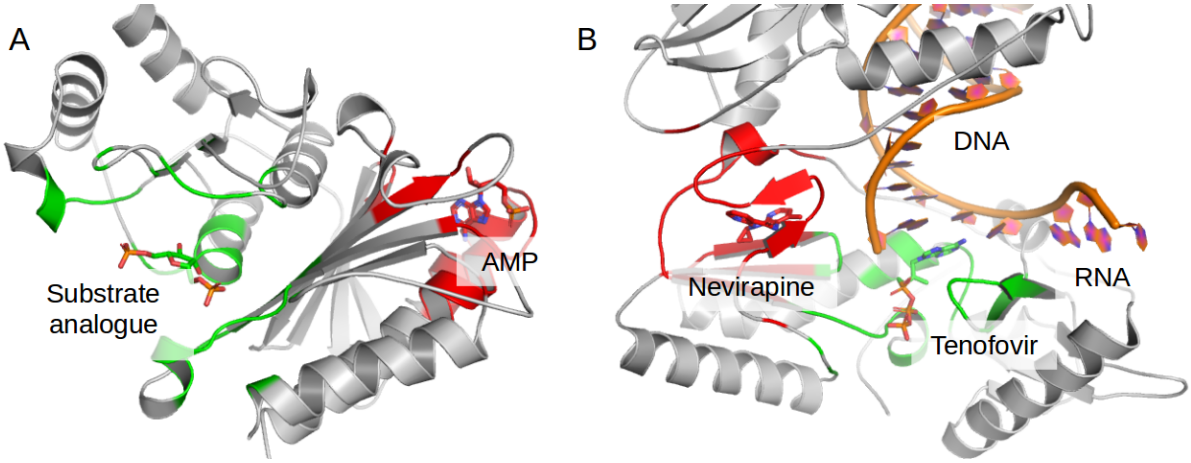


Figure 2-1: Allosteric modulation of enzyme. (A) Fructose 1,6-bisphosphate (PDB: 1FPG), a rate limiting enzyme in gluconeogenesis, bound to a substrate analogue (2,5-anhydro-d-glucitol-1,6-bisphosphate) and the indigenous allosteric inhibitor AMP. Orthosteric sites are colored green and the allosteric sites are colored red. Ligands are shown in stick model. (B) Nevirapine and tenofovir, allosteric and orthosteric inhibitors of HIV-1 reverse transcriptase (PDB: 1T05, 4Q0B).

Druggability. Definitions of 'druggability' vary. Generally, it is referred to the ability of a protein target to bind small molecules with high affinity (Edfeldt et al., 2011). Sometimes, perhaps more appropriately, it is called the 'ligandability'. The druggability score, on the other hand, indicates the probability of such binding. Therefore, the target proteins having more than one binding sites may have different druggability scores for the orthosteric and different allosteric sites. Orally bioavailable drug-like small molecules tend to have properties within certain parameters defined by the Lipinski's rule of five: molecular weight less than 500, logP less than 5, maximum of 10 hydrogen bond acceptors and 5 donors (Lipinski et al., 1997). In order to bind such compounds, a protein should have a binding site with complementary properties e.g. a buried not so polar cleft with appropriate

size to accommodate a drug-like ligand. Druggability score, thus, depends on the pharmacophoric competence between the binding site and the ligand. Structure-based prediction methods rely on identifying cavities in protein crystal structures and assessing the properties of these cavities to predict whether they may bind drug-like molecules. Rules for properties that indicate a druggable cavity are learnt from analysis of co-crystal complexes with drug-like ligands e.g., volume, surface area, polar surface area, surface exposure etc. These rules are then applied to new targets to predict/score druggability. There are several algorithms available to predict the druggability, which includes PocketFinder, Druggability Indices, DoGSiteScorer and DrugEBIlity (An et al., 2005; Hajduk et al., 2005; Volkamer et al., 2012). However, the druggability can also be described in terms of the types of ligand. A protein or a binding site may be more druggable for a particular type of ligands and less druggable for another type. Similarity, allosteric and orthosteric sites on a protein often show differential preference for ligands. Mixing of different types of ligands in the dataset, in turn, produces noise in the large scale screening against a druggable protein. Therefore, in order to reduce entropy and streamline the lead discovery process, it is necessary to develop a method to filter orthosteric and allosteric ligands in real time in a virtual screening setup.

In this chapter we discussed a docking based method that combines the allostery detection and ligand specific druggability scoring within the virtual screening setup. This method enables us to segregate orthosteric and allosteric ligands from a mixed dataset for lead discovery in drug design. Our approach also provides a smart way to depict and compare the molecular docking results. We used genetic algorithm for docking of ligands on the target protein. Blind docking approach was taken and the whole protein was placed in the search space to probe all possible binding sites present in the target. Relying on the stochastic property of docking we generated enough docked conformers to populated all the druggable sites in the target protein. State of the bound conformers in the ensemble was then transcribed into two dimensional density maps (patterns). High density clusters were observed for lead like molecules for the druggable sites on the target. Such, novel pattern based analysis of the clusters provides a way to

probe new druggable sites on a target protein in a much simpler setup. More diverse the ligand library is, more probable the allosteric lead detection becomes. This structure based analysis method would be useful to develop allosteric modulators and to identify novel target sites on drug resistant proteins.

MATERIALS AND METHODS

Similar to the specificity filter as discussed in the previous chapter, this docking based method to find the binding sites and their druggability indices for a test ligand, also requires an ensemble of docked conformers to be analyzed. In order to generate a statistical ensemble of bound conformers for a test ligand, AutoDock 4.2 of The Scripps Research Institute was used as the base program in this method as well (Morris et al., 2009). The structures of the proteins were obtained from the Protein Data Bank (Berman et al., 2000)and the ligand molecules were either obtained from PubChem (Kim et al., 2016) or drawn on Avogadro (Hanwell et al., 2012) followed by geometry optimization in molecular mechanics force filed using the steepest descent algorithm. The center of mass for all the ligands in the library were set to (0,0,0) and the target protein was placed some distance away (around 40 Å) along a Cartesian axis in order to avoid any overlaps. Moreover, the protein was oriented in such as way so that it breaks any symmetry in this overall arrangement. The ligand and the target was place in such a way in order to get a better resolution in the output. Proteins and the ligands were preprocessed in AutoDockTools (Morris et al., 2009) for docking, where the non-polar hydrogens were merged, Gaisteiger charges were added and the atom types were defined. Any water, duplicate atoms or undesirable chemical entities were also removed from the protein structure files. Rotatable bonds in the ligands were defined for the flexible docking. Search spaces were defined by encompassing the whole protein, which is essential in order to explore all the possible binding sites on the target. Lamarckian genetic algorithm with Solis & Wets' local search methods was used as the search algorithm (Morris et al., 1998). Genetic algorithm runs were set to 100 or 1000 with 25 millions energy evaluations each time. Spatial deviation of

the conformers and their binding energies were then distributed on a smooth density histogram in Wolfram Mathematica. Intensity of the densities on these histograms corresponds to the druggabiligy of a site by the ligands. Comparison of densities across a population of ligands contributes to computation of the global druggability score of the site or the protein as a whole.

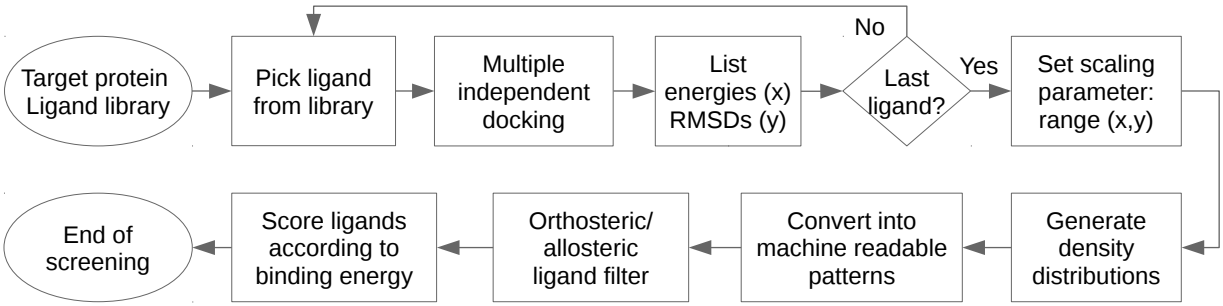


Figure 2-2: The algorithm for the detection and segregation of orthosteric and allosteric leads in a virtual screening setup.

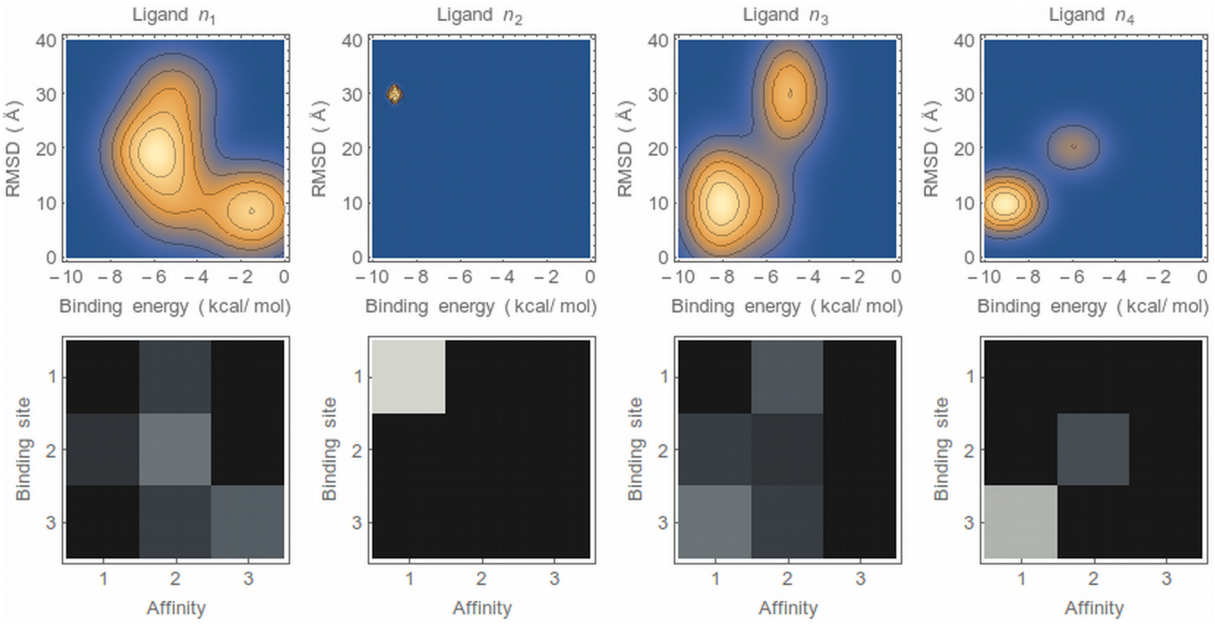


Figure 2-3: Machine readable pattern generation from the density distributions. Upper panel shows the representative distributions. Lower panel shows the patterns generated from the distributions. Each of the distributions are associated with a particular ligand in the dataset describing their binding site preference on the target and affinity for that site.

The scheme of the protocol is shown in the figure 2-2. According to this

protocol a ligand is picked up from a library of small molecules and then docked onto the target protein. The docking runs in a loop until a specified number of iterations is reached. Each docking calculation in this loop is independent of the previous ones. After a certain number of docked conformations are generated, their binding energy, spatial orientation and position are logged. Ranges for the distributions are then defined and the system states are distributed over the energy and RMSD axes on two dimensional smoothed density histograms. Density scaling is performed to make the distributions comparable. The smoothed density histograms are then converted into machine readable patterns (Figure 2-3). Each of these patterns are associated with a particular ligand in the dataset. The patterns describe their binding site preference on the target and affinity for that site. The different types of ligands are then filtered using a segregation function and scored according to the druggability of the target protein/sites toward the ligands.

RESULTS AND DISCUSSION

In very large scale analysis of drug like molecules, virtual screening often suffer from noise (wrong predictions); one possible cause for which, is mixing of orthosteric and allosteric ligands. Here we described the development of a docking based method under virtual screening setup to explore all the druggable binding sites on a target protein for a set of test ligands in order to segregate the orthosteric and allosteric hits. This method also relies on the stochastic nature of docking and pharmacophoric properties of the binding sites as in the specificity calculations described in the previous chapter.

Figure 2-4 shows the interaction of ISO-1 (4,5-dihydro-3-(4-hydroxyphenyl)-5-isoxazoleacetic acid methyl ester) with human macrophage migration inhibitory facctor (MIF). ISO-1 is a known inhibitor of human MIF. It binds to the active site of the enzyme and competitively inhibits the tautomerase activity (Lubetsky et al., 2002). Details of ISO-1 and MIF binding interaction are known from the crystal structure of the complex (PDB: 1LJT). Human MIF is a target protein for the

development of anti-inflammatory drugs. MIF is a homotrimeric enzyme that catalyzes the conversion of Keto-phenylpyruvate to enol-phenylpyruvate and L-dopachrome to 5,6-dihydroxyindole-2-carboxylate. The homotrimeric enzyme shows a three fold symmetry and there are three active sites are located near the N-terminal regions in between two adjacent subunits of the enzyme. The binding of ISO-1 with human MIF was analyzed using the protocol described in Figure 2-2 and compared with the known crystal structure (PDB: 1LJT). The ligand was separated from the complex and placed in the origin (0,0,0); the receptor was moved to around 40 angstroms away from the ligand. 100 docked conformers were generated from independent docking simulations and is shown superimposed in Figure 2-4A. It was found that ISO-1 docked exclusively into the active sties of the enzyme. Then, we generated the probability pattern from this docking experiment and it is shown in the Figure 2-4B. Three high probability densities can be observed in this two dimensional diagram, each of which corresponds to the binding to one of the three active sites. Probability densities are also shown in a 3D histogram representation in Figure 2-4C. When the absolute positions of the ligand and the target are kept fixed, each of such densities along the RMSD axis of the density diagrams represents a separate binding site. In this particular example, all the three densities indicates the orthosteric sites due to the multimeric nature of the protein. Binding to allosteric sites also produces distinct densities. Therefore, using machine learning techniques allosteric and orthosteric ligands can be filtered by analyzing their docked patterns.

4-((3-(4-Hydroxy-3-methoxyphenyl)-4,5-dihydroisoxazol-5-yl) methoxy) benzaldehyde (CP4b) is another reported inhibitor of human MIF (Alam et al., 2011). The docking pattern of this molecule on MIF was also analyzed and compared to that of ISO-1 (Figure 2-4D to 2-4F). All the docked conformers are superimposed and shown in figure 2-4D and the figure 2-4E shows the density patterns. It was observed that, unlike ISO-1 binding, the binding of CP4b was not exclusive to the active sites (Figure 2-4D). Three densities were observed in the density diagram (Figure 2-4E) corresponding to the binding into the active sites and these were comparable energetically to that of ISO-1. However, the probabilities of binding of

CP4b into the active sites of MIF were very low (Figure 2-4F). Other densities corresponding to the binding into other sites of MIF was observed (Figure 2-4E and 2-4F). Although CP4b is an active compound against human MIF, from the analysis of docked patterns it appears that the CP4b binding lacks specificity of interaction. Therefore, this pattern analysis technique also provides us a way to study the specificity of interactions. More specific the interaction is, more dense the cluster becomes. The spread of the cluster inversely correlates with the specificity of interaction as well. Highly specific interactions produce localized and solitary high density spots, whereas, non-specific interactions produce multiple very low density spots spreaded all over the density map.

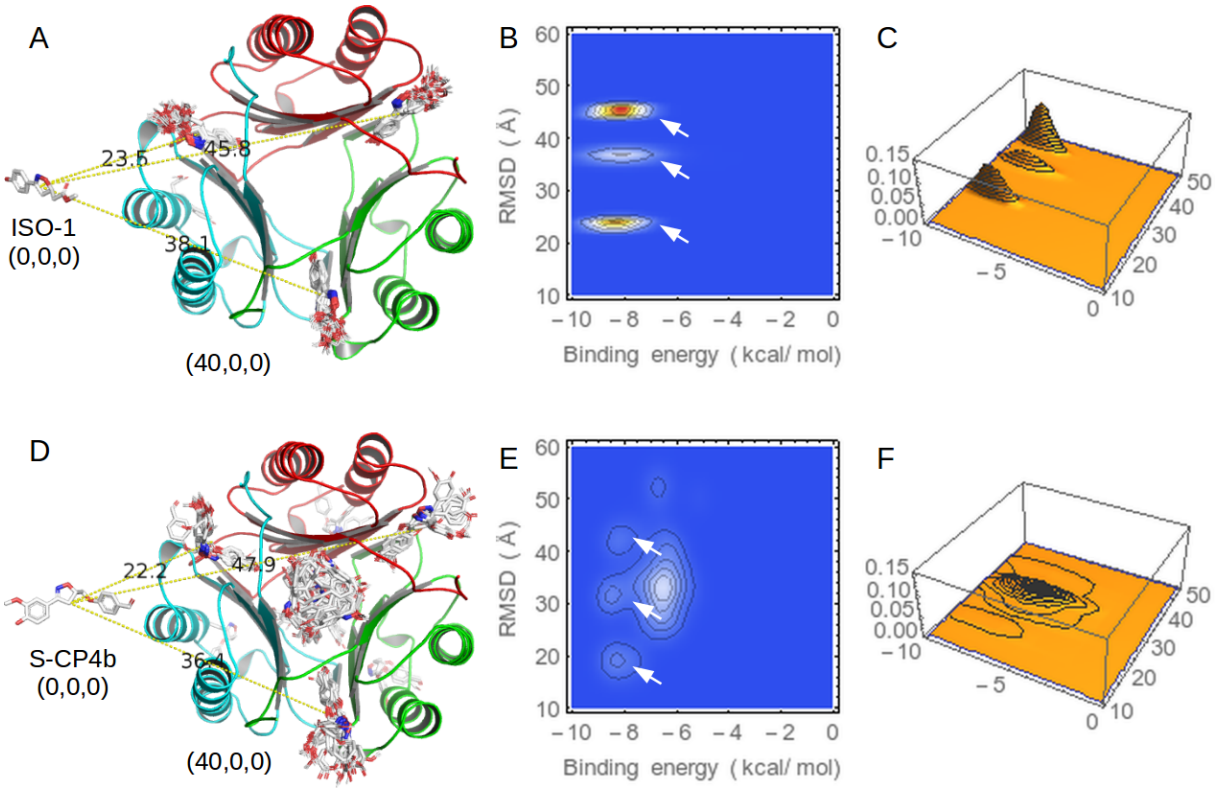


Figure 2-4: Orthosteric ligand binding into the active site. (A) ISO1 binds to the active sites located in between the adjacent subunits of human MIF. (B) Densities of the docked confomers of ISO-1 on the energy-RMSD surface. (C) 3D histogram of the probability densities. (D) Binding of CP4b with human MIF. (E) Densities of the docked conformers of CP4b. Densities in the active sites are marked with arrows. (F) 3D histogram of the probability densities for CP4b binding with MIF.

Figure 2-5 shows the virtual screening of six ligands (epoxyazadiradione, azadiradione, azadirachtin A, azadirachtin B, nimbin and salannin) against a target

protein (human MIF). These compounds are natural products isolated from *Azadirachta indica*. One of the compound, epoxyazadiradione, is reported to be a potent inhibitor of human MIF (Alam et al., 2012). Binding of these compound was analyzed following the protocol given in figure 2-2. It was observed that these molecules produces densities around 29 angstroms along the RMSD axis. This density does not correspond to the active site of the protein but an allosteric site located on the N-terminal side along the C_3 symmetry axis of the protein. Epoxyazadiradione produced the most condensed spot suggesting it to be a highly specific allosteric inhibitor of MIF. Azadiradione also produced a condensed spot, however, it was energetically less favorable. All the other compounds were less specific or energetically less favorable. However, all these molecules showed affinity toward an allosteric site on the target. The origin of such affinity is in their structural similarity. Our method, therefore, can identify and sort the allosteric ligands from a mixed dataset and reduce the noise in the virtual screening.

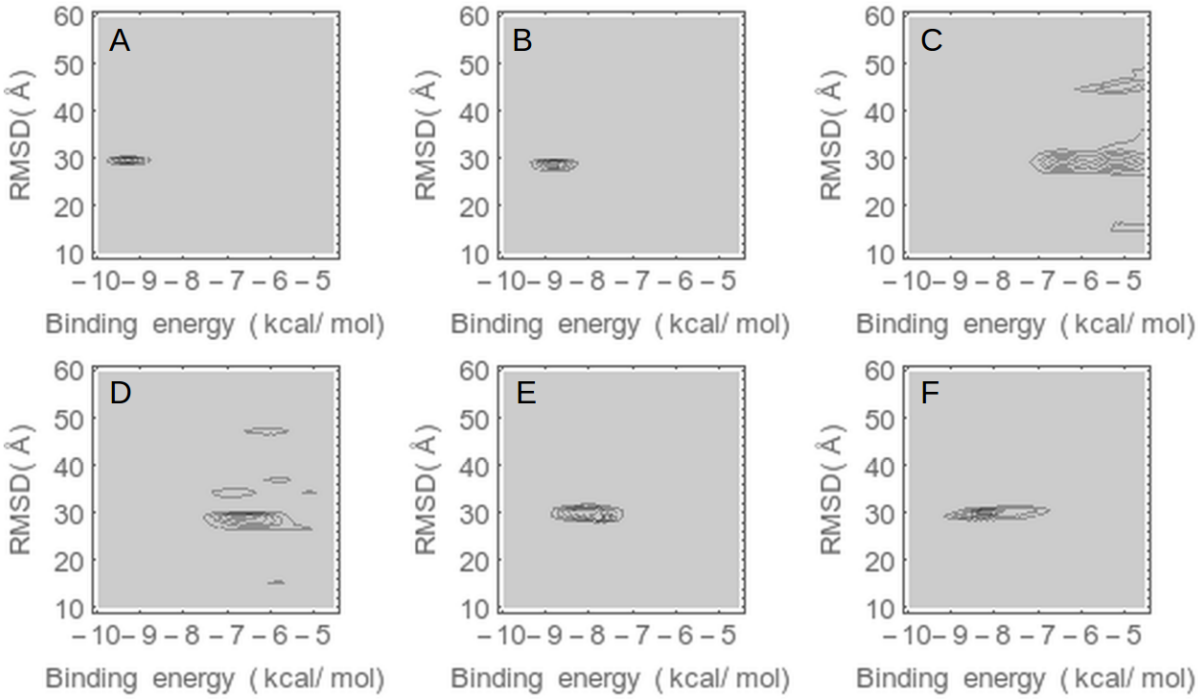


Figure 2-5: Screening of six ligands against human MIF. (A-E) Epoxyazadiradione, azadiradione, azadirachtin A, azadirachtin B, nimbin and salannin binding to human MIF, respectively.

It was observed that not all the sites are druggable by all the ligands. Binding of ISO-1 was exclusive to the active site, whereas, epoxyazadiradione targeted

exclusively the allosteric site. Therefore, in order to probe all the possible binding sites on a target protein a heterogeneous set of ligands is required to be docked, which eventually gets sorted during the analysis. In our algorithm, we used RMSD as a parameter to distribute the system states, however, other parameters such the center of mass or center of geometry etc. can also be used, instead. The method separates the different types of ligands and the binding modes on two dimensions as in the 2D gel electrophoresis, but unlike electrophoresis, here we can go beyond the two dimensions. Several other properties of the ligands such as the change in the solvent accessible surface area etc. can also be included as other dimensions.

CONCLUSIONS

In this chapter we presented a visually attractive way of presenting and analyzing molecular docking results. Such visualization provided us insights into the nature of ligand binding and lead to the development of a new algorithm to probe the allosteric binding sites on a target protein and to reduce noise in virtual screening by segregating the different types of ligands in a mixed dataset. Visualization has become an important scientific tool, especially in the analysis of complex situations. Here, we described how to spread the surface of a protein into a topological map, on which the orthosteric and the different allosteric sites appear as probability densities for ligand binding. Using this method allosteric sites on the target protein can be probed and the druggability of such sites can be determined.

CHAPTER 3 | FINDING THE BINDING SITES IN DISORDERED PROTEINS: A SEQUENCE ANALYSIS APPROACH

Keywords: IDPs, proteome, phylogenetic tree, ANCHOR, GRAVY, pI, amino acid composition, secondary structure

INTRODUCTION

Recent investigations and genome analysis revealed the unique presence of intrinsically disordered proteins (IDPs) in eukaryotes (Dunker et al., 2000; Monsellier et al., 2008; Xie et al., 2007a) and more than 30% of amino acid residues in human proteome are believed to be in the disordered regions of proteins (Dosztányi et al., 2010; Habchi et al., 2014; van der Lee et al., 2014; Peng et al., 2014). High content of disorderedness in proteome suggests a functional role of such regions (Cumberworth et al., 2013; Fuxreiter et al., 2014; Haynes et al., 2006). The presence of disorder regions in a protein is thought to confer large plasticity to interact efficiently with several targets, as compared with a globular protein with limited conformational flexibility (Dunker et al., 2005; Romero et al., 2001; Wright and Dyson, 1999). Thus, the disorderedness of proteins is believed to play significant roles in several biochemical processes, and have been linked to various molecular recognition processes such as DNA binding, cell cycle regulation, membrane transport and other important cellular functions (Dunker et al., 2002; Dyson and Wright, 2005; Tompa and Csermely, 2004; Xie et al., 2007a). The disorderedness, thus, became an intense topic to know its composition, genomic distribution, cellular localization and energetic aspects linked to function and binding to a targeted partner-molecule.

IDPs lack a compact well defined three dimensional structures in their native state and may instead have a number of thermodynamically stable inter-converting

states (Babu et al., 2011; Edwards et al., 2009; Orosz and Ovádi, 2011; Uversky et al., 2000; Vucetic et al., 2003). Some of these proteins are completely unfolded and some contain both the disordered and folded domains with the degree of disorderedness varying from protein to protein (Chen et al., 2006; Dunker et al., 2000). These proteins also have no consistency in their sizes and structurally they resemble the denatured states of ordered proteins (Ahmad et al., 2005; Huang et al., 2006; Uversky, 2002; Weinreb et al., 1996). In a solution, even under physiological conditions, these proteins exist as flexible ensembles of rapidly inter-convertible native conformations (Ahmad et al., 2005; Cohlberg et al., 2002; Huang et al., 2006; Uversky, 2002; Uversky et al., 2000; Weinreb et al., 1996). The binding of a disordered protein to a target molecule or its interaction-partner often causes folding and structural transformation, particularly when it binds to a structured partner/protein. Figure 3-1 shows an example of the structural adaptability of a disordered protein α-synuclein under certain conditions. α-Synuclein, which remains in completely coil conformation in aqueous buffer attains predominantly alpha helical structure upon binding the membrane (Figure 3-1A). Figure 3-1A also highlights the flexible ensembles of rapidly inter-convertible native conformations of α-synuclein solved by NMR spectroscopy. Figure 3-1B shows how the binding to a target protein induces beta sheet structure in α-synuclein. Under pathological conditions, intermolecular interactions may even induce beta sheet structure leading to the aggregation of α-synuclein (Figure 3-1C). Structural alterations of the binding region may, therefore, render a protein function more effectively and with certain specificity (Dyson and Wright, 2002). Recent studies indicated that binding of the disordered proteins precedes global folding and the interactions follow a complex energy landscape. Conformational transformations in a disordered region are often much larger than the changes in globular proteins (Bordelon et al., 2004; Dunker et al., 2002; Gunasekaran et al., 2003; Sugase et al., 2007). As such, the IDPs can bind to several different partners or interfaces and perform diverse functions (Sharma et al., 2014).

In human proteome and other species, a significant number of proteins are found, which are involved in cellular activity but lack any globular fold in their

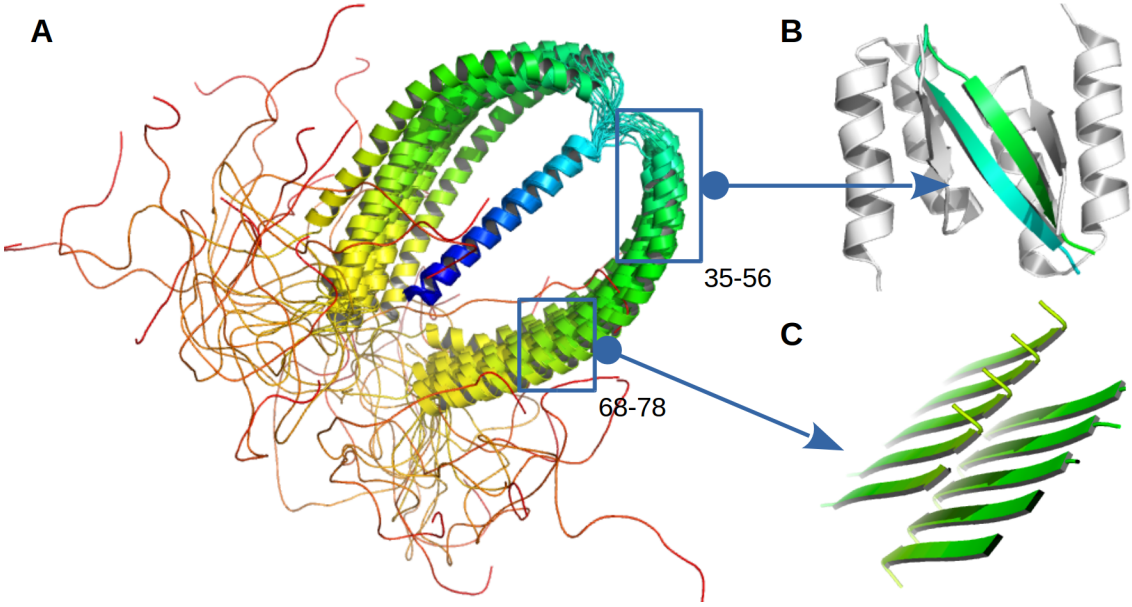


Figure 3-1: Structure of a disordered protein, α-synuclein (UniProtKB: P37840). (A) Micelle bound α-synuclein solved by NMR spectroscopy shows predominantly alpha-helical structure (PDB: 2KKW). An ensemble of thermodynamically stable native conformations are shown. N-terminal to C-terminal of the protein is colored in rainbow (violet to red). (B) Binding with a protein induced beta sheet structure in α-synuclein (PDB: 4BXL). (C) α-synuclein takes beta sheet structure when self aggregates to form insoluble fibrils (PDB: 4ZNN, 4RIL).

native state. It represents at least 30% of human proteome and they play a seminal role in cell signaling, memory storage and other cellular function (Nguyen Ba et al., 2012). Therefore, it is important to understand the differences in the structure-function paradigm as it applies to globular proteins and to IDPs. Apart from functional role, numerous IDPs are associated with several human diseases, including cancer, cardiovascular disease, amyloidosis, neurodegenerative disorder and diabetes (Babu et al., 2011; Xie et al., 2007b). Some of these diseases are inextricably attached to IDPs. α-Synuclein, tau, amyloid beta (Aβ) are among many IDPs involved in diseases like Parkinson's disease (PD) and Alzheimer's disease (AD). Also, it was observed that α-synuclein and tau binding often lead and accelerate the aggregation of the proteins and formation of amyloid. Understanding intermolecular interaction or binding among IDPs and other candidates such as small organic molecules and small peptides, therefore, is an interesting area to explore. The IDPs, as such, could be attractive targets for designing drug molecules that may modulate the protein-protein interactions (Apetri et al., 2006).

Some regions of the IDPs are prone to interact with target molecules and act as binding regions (BRs) or functional part of the proteins. Short functional regions in the disordered region, based on computational analysis and other prediction methods, were detected and termed as molecular recognition features (MoRFs) (Cheng et al., 2007; Disfani et al., 2012; Mohan et al., 2006; Vacic et al., 2007). Computational studies and experimental investigations further verified that BRs in IDPs are exposed and often considered as a primary contact site for the interaction and binding (Csizmók et al., 2005). These regions frequently showed structural propensities similar to the structure they attained upon complex formation with the partner molecule (Dancheck et al., 2008; Fuxreiter et al., 2004). In the present investigation, we aimed to derive the statistical parameters of the physicochemical properties linked to BRs in the intrinsically disordered human proteome. Elucidating binding regions (BRs) and associated statistical knowledge is crucial to address functional and binding roles of the proteins. Our result provided the models of statistical distributions on different aspects of the BRs in IDPs such as the occurrence of BRs, their length, percent occupancy in the parent proteins and the correlations with the degree of disorderedness of the proteins.

We selected all the experimentally validated and annotated proteins with different degrees of disorderedness from IDEAL (Intrinsically Disordered proteins with Extensive Annotations and Literature; release 21 March 2014) and DisProt databases (release 6.02) (Fukuchi et al., 2012; Sickmeier et al., 2007). Several computational methods are available to predict the binding region in disordered protein region (Sharma et al., 2014) such as MoRFpred, DISOPRED3 and ANCHOR. Among them, MoRFpred and DISOPRED3 are developed to predict short protein-binding regions in disorder region which are implicated in molecular recognition processes (Jones and Cozzetto, 2015). We used ANCHOR method to detect the BRs in disordered protein dataset. Short segments of unfolded proteins that showed propensity to interact with some target molecules (mainly proteins) with a possibility of structural recognition are key in detection of binding region by ANCHOR method (Dosztányi et al., 2009; Mészáros et al., 2009). The method utilizes a statistical potential matrix based on pairwise interaction energy from

known coordinates using a dataset of globular proteins (Dosztányi et al., 2005, 2009; Mészáros et al., 2009). ANCHOR is independent of amino acid composition, although it was reported that the construction of the algorithm for the prediction of interaction energy implies its sensitivity to amino acid composition (Mészáros et al., 2009). We found more than 3000 binding regions in the human disordered proteome which were partly or fully disordered in their native state.

Different statistical models were invoked to describe the distribution pattern of frequency and length of the BRs, hydrophobicity and many other properties which are very crucial for the binding regions for their functional activity. The statistical pattern and associated parameter so derived could be used to predict the behavior and property of new proteins that contain certain degree of disorderedness. This information could be useful in medicine and interest in protein disorder as a possible target for designing drug molecules. Partial or fully disorder states play critical role in cell signaling and also linked to several disease formation and therefore the binding regions could be the targets for designing drug molecules to arrest/stop progression of disease linked to protein disorder.

MATERIALS AND METHODS

Compilation of dataset. Information of the human IDPs was obtained from IDEAL (Intrinsically Disordered proteins with Extensive Annotations and Literature; release 21 March 2014) database and DisProt database, release 6.02 (Fukuchi et al., 2012; Sickmeier et al., 2007). We retrieved 471 unique protein sequences from UniProt (release 2014_07) after ID mapping. IDEAL database entries have extensive annotations of disorderedness. DisProt also lists the IDPs detected by experimental methods such as fluorescence, circular dichroism, FTIR, sensitivity to proteolysis etc. Therefore, our dataset comprised experimentally determined and extensively annotated IDPs and represented human disordered proteome. Sequences were obtained from UniProt in FASTA format and then converted to strings of one letter amino acid codes for further analysis.

Sequence comparisons and phylogenetic analysis. Global alignment of the sequences was performed at Clustal Omega, which is a multiple sequence alignment program available at EMBL-EBI web server (http://www.ebi.ac.uk/Tools/msa/clustalo/). It uses seeded guide trees and hidden Markov model profile-profile techniques to generate alignments between sequences. Phylogenetic tree was generated from the sequence alignment using the neighbor-joining algorithm to construct trees from the distance matrix by neighbor joining method in Newick format (Cardona et al., 2008). The tree was rendered at the Interactive Tree Of Life (iTOL) server, which is an online tool for the display and manipulation of phylogenetic trees (http://itol.embl.de/).

Calculation of disorderedness and binding regions. Disorderedness of the proteins was computed using the IUPred program (Dosztányi et al., 2005). The ANCHOR method was engaged to detect the binding regions in the IDPs. ANCHOR analyzed the input sequences of unfolded protein and predicted the binding regions based on certain scoring values (Dosztányi et al., 2009; Mészáros et al., 2012). BR sequences were obtained from the protein sequences using the position and length parameters.

Sequence analysis. Length, amino acid composition, charged residues, total charge, and molecular weight were calculated from the sequence data. The GRAVY value for a BR or protein is calculated as the average of hydropathy values (Kyte and Doolittle, 1982) of all the amino acids. Isoelectric points were calculated using the Compute pI/MW tool (Bjellqvist et al., 1994; Gasteiger et al., 2005) at ExPASy Bioinformatics Resource Portal. Computational algorithm PSIPRED was used to predict the conformation propensity for each protein from their amino acid sequence (Jones, 1999). Larger proteins were segmented into domains using DomPred prior to secondary structure prediction. Percentage of residues in a protein with preference for a particular conformation was measured by taking a ratio of the total number of residues preferring a particular conformation to the protein sequence length. Secondary structure compositions of BRs were obtained from the parent protein analysis using the position and length parameters.

Statistical analysis. All the statistical analysis was performed in Wolfram Mathematica 10. Cramér-von Mises test (Anderson and Darling, 1952) was used to test the normality of the data. For the normally distributed data, mean, standard deviation (SD) and standard error of mean (SEM) were calculated. Significance of the mean differences was established with Student's t-test and the null hypotheses were rejected at the 5 percent level of significance. Probability values of less than 0.0005 were considered as highly significant and denoted by *** in the graphs. Likewise, the probability values in between 0.0005 and 0.005 were considered as very significant and denoted by ** in the graphs and the rest were denoted by a single star. Poisson distribution was fitted to the BR frequency and length data (a discrete random variable). The probability mass function (PMF) is given by:

$$f(x;\mu)=\frac{e^{-\mu}\mu^{x}}{x!}\ , \qquad \text{Eq. 3-1}$$

where e is Euler's number and μ is the expected value of the random variable x. The cumulative distribution function (CDF) is given by:

$$g(x;\mu)=e^{-\mu}\sum_{i=0}^{\lfloor x\rfloor}\frac{\mu^{i}}{i!}\ , \qquad \text{Eq. 3-2}$$

where ⌊x⌋ is the floor function. Generalized Poisson distribution function or the Poisson-Consul distribution is given by the equation 3.

$$f(x;\mu)=\frac{e^{-x\lambda-\mu}(x\lambda+\mu)^{-1+x}}{x!}\ , \qquad \text{Eq. 3-3}$$

where λ is any real number between 0 and 1. Normal distribution with mean (μ) and standard deviation (σ) were fitted to the normally distributed data. The probability distribution function is given by:

$$P(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^{2}}{2\sigma^{2}}}\ , \qquad \text{Eq. 3-4}$$

The cumulative distribution function is given by:

$$D(x)=\frac{1}{2}\left[1+erf\left[\frac{x-\mu}{\sqrt{2}\sigma}\right]\right], \qquad \text{Eq. 3-5}$$

where erf is the error function. PDF and CDF of the skewed normal distribution were described by the following two equations, respectively.

$$P(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad \text{Eq. 3-6}$$

$$D(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}Erfc\left[\frac{-\alpha(x-\mu)}{\sigma\sqrt{2\pi}}\right], \quad \text{Eq. 3-7}$$

For the bimodal distributions, a mixture of distributions was fitted to the data. In regression analysis straight lines passing through the origin were fitted to the data.

RESULTS AND DISCUSSION

The protein dataset (*vide* Appendix I) comprised experimentally determined disordered proteins obtained from IDEAL and DisProt databases. These proteins were extensively annotated IDPs and represented human disordered proteome. Based on the content of structural disorderedness, the proteins were grouped into three categories as suggested in previous reports (Das et al., 2014; Schad et al., 2013). In the dataset, the total number of largely disordered proteins (LDP, structural disorderedness >70%) was 71. 163 proteins with disorderedness ranging from 30 to 70% were grouped as moderately disordered proteins (MDP). Rest of the proteins, having less than 30% disorderedness, was grouped as partially disordered protein (PDP).

We compared all the sequences of human disordered proteome in the dataset using the multiple sequence alignment tool, Clustal Omega. A phylogenetic tree was derived from this alignment to study the similarity and evolutionary distances between the sequences. Closely similar proteins were grouped together and the dissimilar proteins got separated in different groups. The relationships among the proteins are shown in cladograms (Figure 3-2). The detailed tree with branch length information and leaf labels is freely available via Internet at http://jpp.org.in (Pal et al., 2016). Figure 3-2 shows the cladograms of the tree with 471 leafs in the rooted and unrooted modes. Some important disease related disordered proteins such as α-synuclein, BRCA1, p53 and amyloid-β are marked. The

tree suggests that the α-synuclein is closer to BRCA1 in sequence similarity than p53 or amyloid-β. Proteins with different degrees of disorderedness were also color coded to show their distribution among the different clades of disordered proteins. It was found that all the clades have proteins with varying degree of disorderedness.

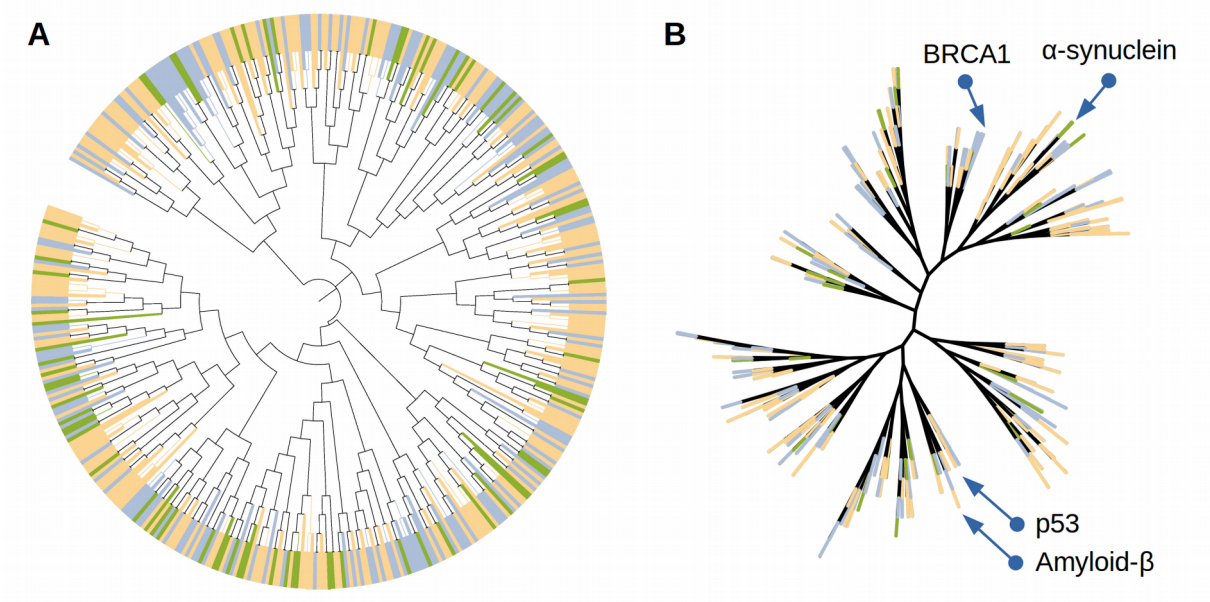


Figure 3-2: Human disordered proteome tree. (A) A 471 leaf tree with colored clades rendered in circular mode. The tree is shown without branch length information. (B) The same tree is shown in unrooted mode. Some important disease related disordered proteins are marked. Color strip dataset was used to define branch colors for different group of disordered proteins: partially disordered proteins (PDP), gold; moderately disordered proteins (MDP), blue; largely disordered proteins (LDP), green.

We used ANCHOR algorithm to detect the BRs in disordered protein dataset. ANCHOR largely depends on the pair wise energy estimation method that is used in IUpred algorithm. IUpred was used to quantitate the disorderedness of the proteins. Binding regions were selected by ANCHOR algorithm by identifying region in the polypeptide chain that are in disordered regions and not supported by favorable intra chain interactions to attain folded structure. ANCHOR detected 3494 binding regions (BRs) 471 unique human proteins with different degree of disorderedness. The detailed lists of the number of BRs in each protein, their total lengths, percent occupancy in the parent protein, sequence compositions,

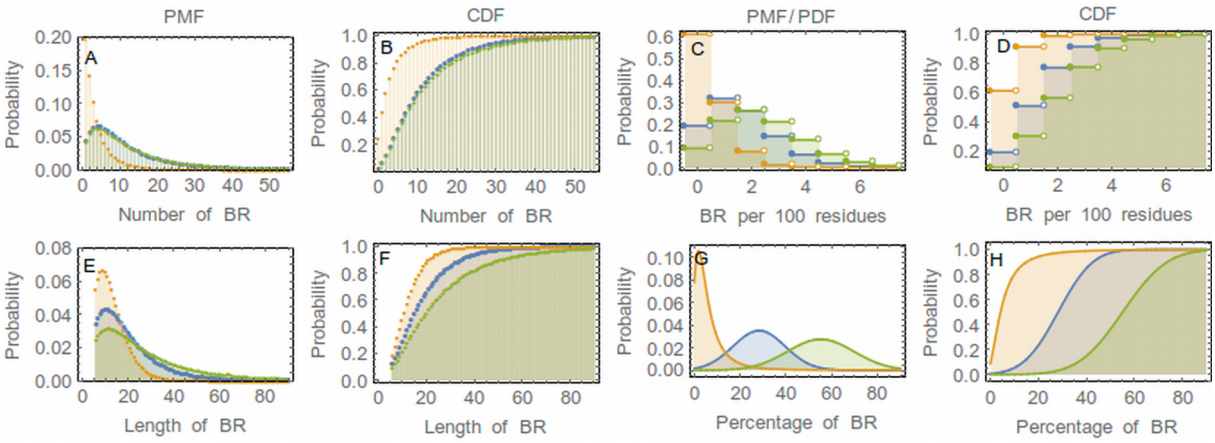


Figure 3-3: Frequency and length distribution of binding regions (BRs). (A) Probability of occurrence of a BR (BR frequency) in different group of disordered proteins. Probability mass function (PMF) of the fitted Generalized Poisson Distribution is shown. (B) Cumulative distribution function (CDF) of the BR frequency. (C) Probability of occurrence of a BR per 100 residues of a protein. (D) CDF of the BR frequency per 100 residues of a protein. (E) Probability distribution (PMF) of individual BR lengths in different group of disordered proteins. (F) CDF of the BR length distribution. (G) The distribution of BR content (percent occupancy) in a protein. PDF of the fitted skewed/normal distribution is shown. (H) CDF of the BR content distribution. Color key: gold, partially disordered proteins (PDP); blue, moderately disordered proteins (MDP); green, largely disordered proteins (LDP).

theoretical physicochemical properties and structural propensities of individual binding regions and the IDPs, are freely available via the Internet at http://jpp.org.in (Pal et al., 2016). Most of the proteins contained multiple binding regions. Figure 3-3A and 3-3B shows the probability distributions of BR frequency in all the three groups of protein, whereas, figures 3-3C and 3-3D show the probability distribution of finding a BR per 100 residues of a protein. Interestingly, the number of BRs did not always follow a normal distribution. Instead, it shows a Poisson distribution pattern suggesting that the occurrence of BRs in a protein is a stochastic process, which satisfies the Markov property (Durbin, 1998; Nguyen Ba et al., 2012). Figure 3-3 shows the fitted poison distributions for BRs in whole protein and also with respect to 100 residues in each group of proteins. The Poisson distribution analysis provided the expectation values (μ), which represent the occurrence rate of the event (here number of BRs). The expected value of BRs was 3 in MDP and LDP. In PDP the expected BR frequency was 1 (Table 3-1). However, the expected values of BRs per 100 residues of a protein were found to

be 0, 1 and 2 for PDP, MDP and LDP, respectively. Interestingly, we observed that the percentage of residues in BR followed a normal distribution pattern. In PDP the normal distribution was positively skewed. A shift in the modal class with the increasing degree of disorderedness in proteins was observed. In the LDPs the percentage of residues in BRs were very high compared to the MDPs or PDPs. The increase in the number of expectation values with protein disorderedness was within the scope of ANCHOR algorithm, which was used to detect the BRs in the disordered protein dataset, however, we described here the detailed statistics of BR frequencies, provided the quantitative parameters and showed how the BR frequencies correlates with the disorderedness of the protein, which would be useful for understanding the origin of binding regions in disordered as well as ordered proteins.

Table 3-1: Statistics of BR distribution. Poisson/Poisson-Consul distribution parameters μ and/or λ for BR frequency, BR frequency per 100 residues of protein and the length of the BRs.

Variables	Parameters	PDP	MDP	LDP
BR frequency	μ	1.41	3.82	3.83
	λ	0.56	0.67	0.70
BR frequency per 100 residues†	μ	0.49	1.65	2.42
BR length	μ	6.40	6.38	6.44
	λ	0.49	0.66	0.75

† Poisson distribution with parameter μ was fitted to the data

We further studied the distribution pattern of the length of BRs with respect to degree of protein disorderedness. We observed that the content of BRs did not follow a normal distribution. Generalized Poison distribution formula fitted the data much better than the normal distribution. Figures 3-3E and 3-3F display the length distribution of the individual BRs and figures 3-3G and 3-3H show the percent occupancy in different group of proteins. We observed an expectation value of the BR length of 6 residues in all the three groups of protein (Table 3-1). However, the spread of the distributions increased with the increase in disorderedness.

In order to better understand how the BR frequencies, BR frequency per 100 residues of a protein, BR length and percent occupancy correlates with the protein disorderedness we performed regression analysis as shown in the figure 3-4. Frequency versus disorderedness plots suggested linear regressions. Therefore, straight lines passing through origins were fitted to the data. BR expectancy per 100 residues produced discrete densities over the continuous axis of protein disorderedness, which gave a much clearer understanding about the expected number of BRs per 100 residues changes depending on the protein disorderedness. Analysis showed that the length of the individual BRs did not correlate with the disorderedness of the protein, which was evident from the very low R^2 value of the linear model fit (Figure 3-4F). However, the percent occupancy of the BRs in a protein increased linearly along with the protein disorderedness with a coefficient of 0.62.

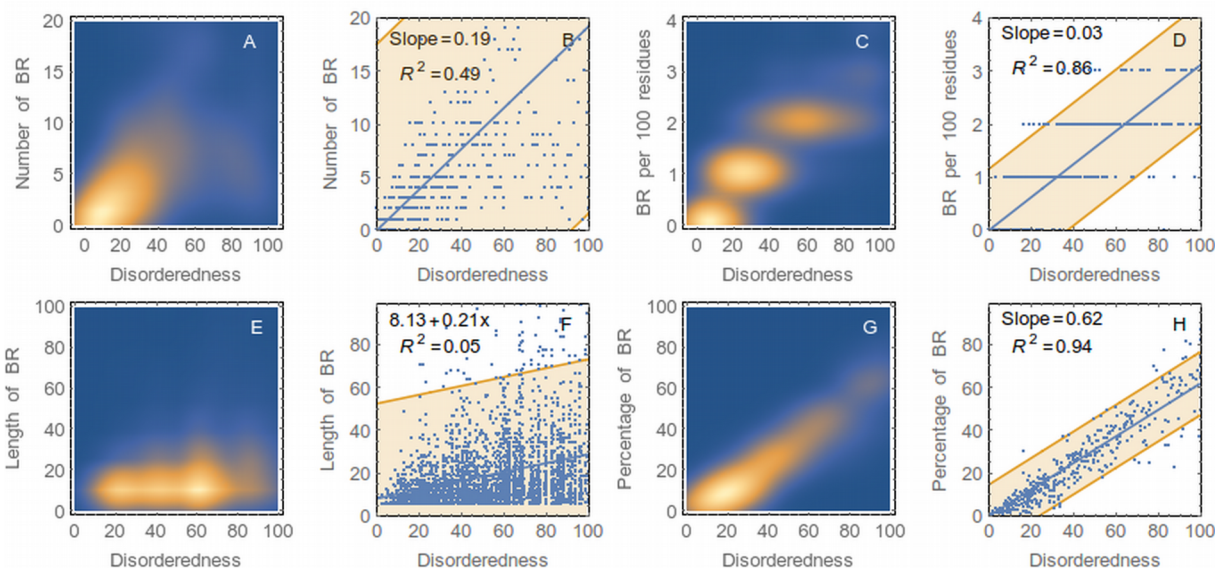


Figure 3-4: Correlation of BR frequency/occupancy with protein disorderedness. Distribution of BR frequency with protein disorderedness (A) and the fitted linear model (B). Distribution of BR frequency per 100 residues of protein with the protein disorderedness (C) and the fitted linear model (D). Distribution of BR lengths with protein disorderedness (E) and the fitted linear model (F). Distribution of BR occupancy in a protein with the protein disorderedness (G) and the fitted linear model (H). Confidence level bands at 95% are shown.

Using the hydropathy indexes (Wu et al., 2006) of individual amino acids grand average hydropathy (GRAVY) values of the BRs and the parent proteins were

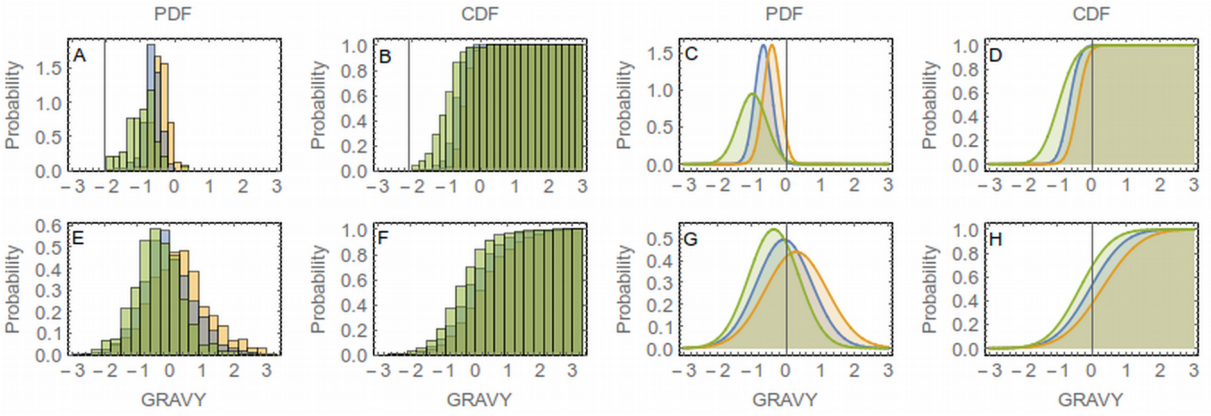


Figure 3-5: GRAVY distribution of the whole proteins versus BRs. (A) Frequency histogram of GRAVY of the IDPs. (B) Cumulative frequency histogram of GRAVY of the IDPs. (C) Fitted normal distribution of GRAVY of the IDPs. (D) CDF of the fitted normal distribution to the GRAVY of the IDPs. (E) Frequency histogram of GRAVY of the BRs. (F) Cumulative frequency histogram of GRAVY of the BRs. (G) Fitted normal distribution of GRAVY of the BRs. (H) CDF of the fitted normal distribution to the GRAVY of the BRs. Color key: gold, PDP; blue, MDP; green, LDP.

derived. The calculated GRAVY indexes of all the proteins were predominantly negative and varied between 0 to -2, approximately (Figure 3-5 and Table 3-2). It was expected as the proteins were rich in polar and charged residues. Mean GRAVY significantly decreased with the increase in the degree of disorderedness (PDP, MDP and LDP) (*vide* Appendix I for the t-test values, Table S3-1). However, the spread of GRAVY values was very high (ranging 2 to -2, approximately) for BRs, mean nearing neutrality (Figure 3-5).

Table 3-2: GRAVY statistics. Fitted Normal Distribution parameters for GRAVY of IDPs and BRs.

Groups	PDP		MDP		LDP	
	μ	σ	μ	σ	μ	σ
Protein	-0.41	0.25	-0.66	0.25	-0.98	0.42
BR	-0.35	0.74	-0.05	0.81	0.30	0.91

Distribution of the isoelectric points (pI) of BRs and the parent proteins are shown in the figure 3-6. Statistical analysis showed that theoretical pI values followed a bimodal distribution for both the proteins and BRs (Figure 3-6 and Table 3-3). pIs were mostly distributed either in acidic or in basic regions, but rarely at the

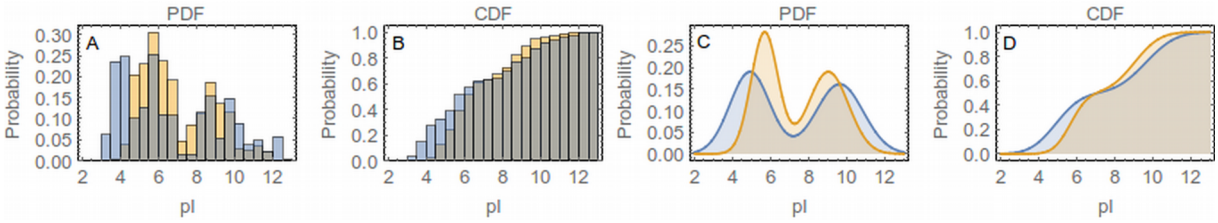


Figure 3-6: Distribution of isoelectric points (pI) in proteins and BRs. (A) Histograms of pI densities. (B) Histograms of cumulative densities of pI. (C) Fitted bimodal distributions. (D) CDF of the fitted distributions to pI densities. Color key: gold, protein; blue, BR.

neutral pH. On both sides, they followed a normal distribution. Such multimodal distributions of pIs for the whole proteome are known in the literature (Kiraga et al., 2007; Nandi et al., 2005; Taylor et al., 2002). pI distribution of the BRs closely followed that of the parent proteins in terms of the density. However, the mean pI of BRs in the acid ranges was found to be significantly less than their parent proteins and in the basic ranges significantly higher.

Table 3-3: Statistics of isoelectric points (pIs). Mean acidic and basic pIs of proteins and BRs. Values are given as μ±σ.

Groups	pI basic	pI acidic
Protein	5.68±0.71	9.03±1.05
BR	4.91±1.05	9.63±1.24

t test (Protein vs BR): pI acidic $3.53295*10^{-76}$; pI basic $1.47157*10^{-29}$

The amino acid composition of the binding regions is shown in figure 3-7 and compared to the protein sequence composition. ANCHOR is independent of amino acid composition, although it was reported that the construction of the algorithm for the prediction of interaction energy implies its sensitivity to amino acid composition (Mészáros et al., 2009). In most disordered regions the functional amino acid residues remain unknown (Nguyen Ba et al., 2012). We have found that the BRs mostly differ from their parent proteins in the content of charged or polar amino acids. Charged amino acids such as Glu, Lys, Arg, and Asp are present in significantly lower amounts in the BRs so are the uncharged polar residues: Thr and Asn. Hydrophobic amino acids such as Leu, Ala, Val, Ile and Phe are more abundant in the BRs. Ser alone in the uncharged-polar group is present in significantly higher

number in the BRs. It is the smallest amino acid in this group having least bulky side-chain. Hydrophobic and hydrogen bonding interactions are the major players in the ligand-protein or protein-protein binding (Jiang et al., 2002). Contribution of electrostatic/ionic interactions is much less compared to them. Therefore, the less abundance of charged and uncharged-polar residues and the prominence of hydrophobic residues in the BRs are desirable and justified. Figure 3-7 also shows the comparison of mean residue molecular weights (MRW) of protein and BRs. Although the mean MRW is similar in both the BRs and Protein, the spread is very high for the BRs.

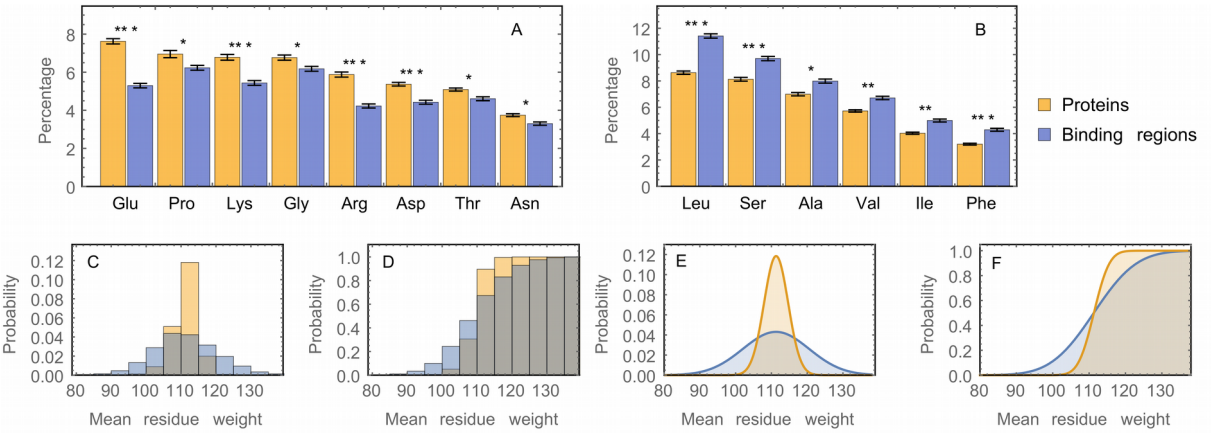


Figure 3-7: Comparison of amino acid composition between whole protein and the BRs. (A) Amino acids that are more abundant in whole protein. (B) Amino acids that are more abundant in BRs. Significant variations were marked with asterisks. ***, p-value <0.0005; **, p-value <0.005 but not <0.0005; *, p-value <0.05 but not <0.005. (C) Comparison of mean residue molecular weight (MRW) distribution of BR and protein. (D) CDF of MRW histogram. (E) Fitted normal distributions of MRW (PDF). (F) Fitted normal distributions of MRW (CDF). Color key: gold, protein; blue, BR.

Elucidating binding regions and associated structure formation on these binding regions in IDPs is significant as this is the starting point for investigations into higher-order structure and, thus, functions of IDPs. The conformations (extended/β-strand, helix and coil), the residues in proteins and in BRs prefer to adopt are shown in figure 3-8. It should be noted that ANCHOR analysis is independent of the adopted secondary structure (Mészáros et al., 2009) and, therefore, the result was not biased by the algorithm. We found that the ordered secondary structure decreased with the increase in disorderedness of the protein.

Coil conformation was found to be the most preferred conformation in all the three groups of disordered proteins, followed by helix and then the extended or β-strand/sheet. BR structural propensity followed a similar trend in all three groups as well. However, in the PDP and MDP groups the BR structure propensity toward coil conformation was significantly higher than that of the proteins in that group. In MDP, the propensity toward helix was significantly lower for BRs. Although the trend is visible, such statistical significance could not be established in other groups. However, when PDP, MDP and LDP data were combined, we observed significantly less propensity toward helix and higher propensity toward coil in the BR residues (Figure 3-8D). Propensity toward extended conformation was also significantly lower in BRs. The overall structural content of BR sequences was: extended ~6%, helix ~18% and coil conformation ~76% indicating that binding region was dominated with sequences that preferred to be flexible. In total protein, however, structural preference of the sequences was: extended 9%, helix ~27% and coil ~64%. The analysis showed that very few residues preferred β-sheet/strand conformation and both the BRs and the parent protein molecules are rich in sequences, most of which preferred coil/random conformation and the propensity for coil conformation is more in BR sequences.

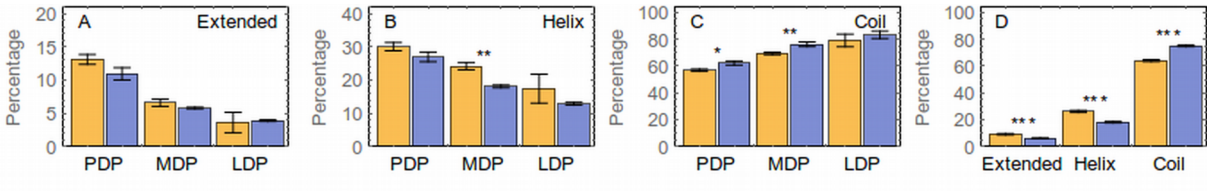


Figure 3-8: Comparison of the secondary structure propensity of the BRs with the proteins. (A) Propensity for extended conformation in three groups of disordered proteins. (B) Propensity for helix and (C) propensity for coil conformation. (D) PDP, MDP and LDP combined. Significant variations were marked with asterisks. ***, p-value <0.0005; **, p-value <0.005 but not <0.0005; *, p-value <0.05 but not <0.005. Color key: gold, protein; blue, BR.

Experimental and computational studies highlighted widespread roles of protein disorder in biological processes (Dunker et al., 2002; Dyson and Wright, 2005; Gsponer et al., 2008; Krishnan et al., 2014; Wright and Dyson, 1999). Recent discovery showed that some protein phase separation leads to formation of

membrane less organelles/component which have important roles in cellular function; IDPs have significant role in the formation of such assembly and localization of many signaling proteins to act efficiently. Protein disorder is also linked with several diseases and, therefore, the disordered proteins are considered important drug targets in rational drug design (Cheng et al., 2006; Lao et al., 2014). However, the disordered protein regions do not act as isolated domains and the surrounding segments in addition to its length also govern its function and stability. Thus, it is very important to characterize the large number of disordered regions along with total protein to realize their greater role in cellular activity and to develop new strategies for drug design targeting specific regions in IDPs.

CONCLUSIONS

Our analysis provided the content, composition and statistical behavior of the binding regions in disordered proteins and some of its physicochemical aspects such as isoelectric point and hydrophobicity. We have shown the distributions of BR lengths and their percent occupancy in the parent proteins. We have described the correlations of BR occurrence frequencies, lengths and percentages with the degree of disorderedness of the parent protein. Statistical models for the occurrence of BRs in disordered proteins were derived. Some parameters followed poison distribution and some others showed normal distribution. Theoretical pI values followed a bimodal distribution. The statistical analysis further illustrated how the linked parameters differed along with the content of protein disorderedness. The report also shows that the BRs contained amino acids optimal for hydrophobic and the hydrogen bonding type of interactions with the target molecule/protein. Hydrophobicity of the BRs was widespread and the pIs were more acidic or more basic than that of the parent proteins. Their structural disposition of BRs towards the more flexible coil conformation was also discussed. It would be interesting to test the binding and functional efficacy of the regions with some of the target molecules.

CHAPTER 4 | PROBING THE MICROENVIRONMENT OF BINDING SITE WITH SMALL MOLECULES: ILLUMINATED WITH FLUORESCENCE

Keywords: serum albumin, new chemical entity, fluorescent probe, hydrophobicity, Stokes' shift

INTRODUCTION

The interaction and the energetics of small molecule binding to a protein largely depend on the molecular architecture and microenvironment provided due to folding/unfolding or even transformation of the protein structure. The observable properties of a small molecule in such microenvironment, in turn, carry information about the binding site, which is crucial for drug development and many other investigation (Abou-Zied and Al-Shihi, 2008; Cohen et al., 2002; Er et al., 2013; Royer, 2006). Our attention was focused on fluorescence emission and binding aspects of the fluorophore in the hydrophobic milieu inside a globular fold of a protein under physiological and denaturing conditions. We have synthesized a naphthalene based fluorophore, methyl 3-[(6-{[2-(tert-butoxy) -2-oxoethyl] (4-methoxyphenyl) amino} naphthalen-2-yl) formamido] propanoate (compound **5**) to study the immediate surroundings of the molecule inside the proteins.

Serum albumin was chosen as model protein with at least seven hydrophobic grooves on its surface. It provides a unique microenvironment and acts as a universal receptor for many drug molecules (Curry et al., 1998; Er et al., 2013; Reichenwallner and Hinderberger, 2013; Simard et al., 2006). This protein increases the solubility of hydrophobic ligands in plasma and modulates their delivery to cells The precise architecture of the binding pockets is known from several crystallographic and NMR spectroscopic studies (Curry et al., 1998; Reichenwallner and Hinderberger, 2013; Simard et al., 2006). Thus, the interaction pattern and the spectroscopic signature of a small molecule housed in the well defined

environment of serum albumin could provide significant insight into the interaction pattern and its binding efficacy (Yamasaki et al., 2013).

Intrinsic protein fluorescence originating from tryptophan and tyrosine residues or the fluorescence of the drug molecule itself provides ample information about the local environment, the changes in protein conformation and the interaction of a protein with a drug molecule (Abou-Zied and Al-Shihi, 2008; Cohen et al., 2002; Royer, 2006). However, the current investigation explored both the quantitative and qualitative aspect of the interaction and incorporation of compound **5** into the binding pockets of serum albumin, at the molecular level using fluorescence methodologies. The thermodynamic parameters were obtained by measuring the effect of temperature on binding constant. In addition to access the interaction site and binding specificity of the drug molecule computational modeling analysis was carried out.

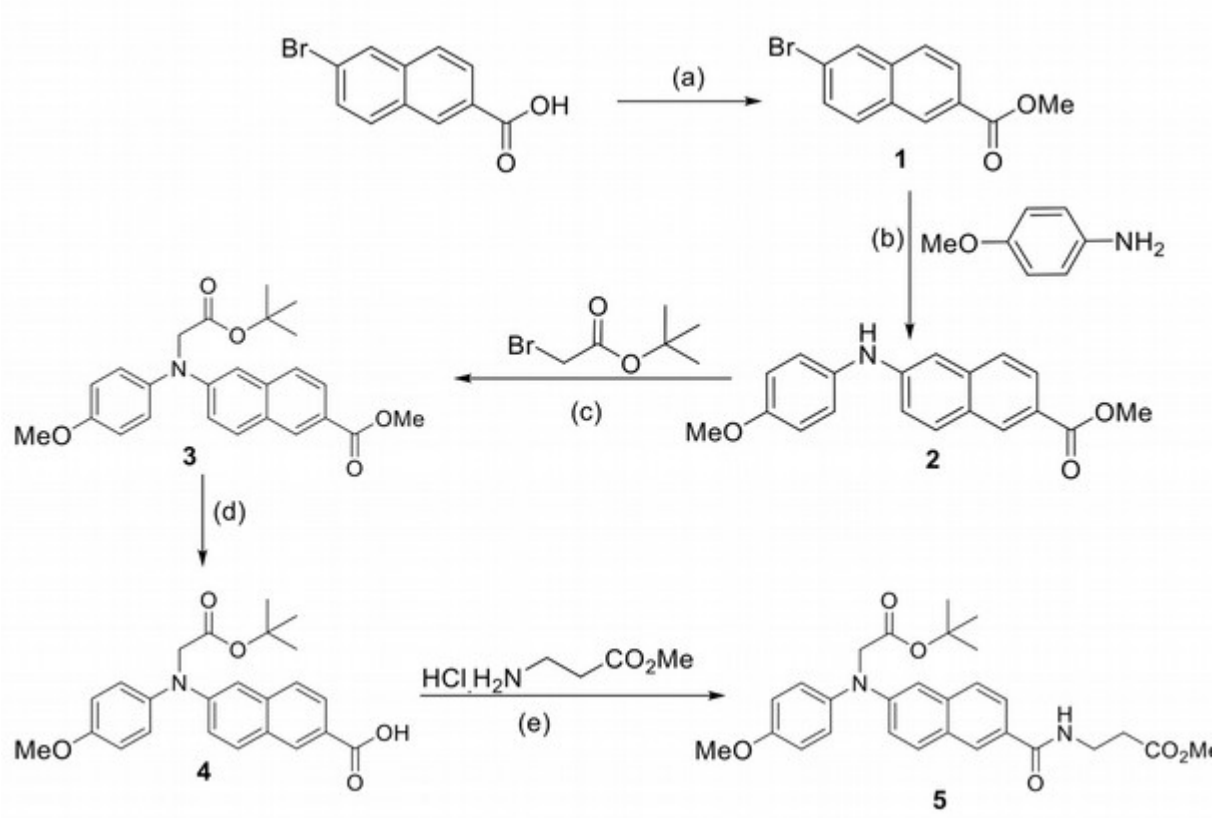
MATERIALS AND METHODS

Chemicals. Bovine and human serum albumins, Tris-HCl and Urea were purchased from Sigma–Aldrich Corporation (St. Louis, MO, USA). All the samples were prepared in 20 mM Tris-HCl buffer of pH 7.0. Deionized and triple distilled water was used for preparing the buffers.

All air and water sensitive reactions were carried out in oven dried glassware under nitrogen atmosphere using standard manifold techniques. All the chemicals were purchased from Acros organics and Sigma-Aldrich, and used without further purification unless otherwise stated. Compounds that are not described in the experimental part were synthesized according to the literature procedures. Solvents were freshly distilled by standard procedures prior to use. Flash chromatography was performed on silica gel (Merck, 100–200 mesh) with the indicated eluant. All ^{1}H and ^{13}C-NMR spectra were recorded on a Bruker 600 MHz spectrometer. For 1H NMR, tetramethylsilane (TMS) served as internal standard (δ

= 0) and data are reported as follows: chemical shift, integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet) and coupling constant(s) in Hz. For ^{13}C NMR, TMS (δ = 0) or CDCl3 (δ = 77.26) was used as internal standard and spectra were obtained with complete proton decoupling. Mass spectra were obtained on a Jeol MS station 700 and ESI-TOF mass spectrometer.

Procedure to synthesize compound 5. Synthesis was carried out following the scheme 4-1. Detailed procedure of synthesis and the characteristic data are given in the Appendix II.



Scheme 4-1: Reagent and conditions: (a) Conc. H_2SO_4, MeOH, 0 °C to r.t., 6 hrs. (b) Palladium (II) acetate (0.05 equiv.), xantphos (0.1 equiv.) and cesium carbonate (3 equiv.), 80 °C, 4 hrs. (c) Potassium tert-butoxide (1.2 equiv.), DMF, 0 °C to r.t., 12 hrs. (d) Lithium hydroxide (3 equiv.), MeOH-water (5:1), r.t., 2 hrs. (e) EDC.HCl (3.0 equiv.), HOBT (2.5 equiv.), TEA (6 equiv.), 0 °C to r.t., 1.5 hrs.

Absorption spectroscopy. Ground-state absorption spectra were recorded with a Shimadzu UV-2401PC Spectrometer. 1 cm path-length quartz cuvette was used and 250-450 nm wavelength range was scanned. Compound **5** absorption

spectra as a function of BSA and HSA concentration were recorded by titrating (Banerjee et al., 2012; Ray et al., 2012) compound **5** solution with concentrated protein solutions. Small dilution error in the titration experiment was ignored.

Fluorescence emission and excitation spectroscopy. The steady-state fluorescence emission and excitation spectra were recorded with a Cary Eclipse Fluorescence Spectrophotometer. The emission spectra of serum albumins and compound **5** were obtained by exciting the samples at the wavelengths 295 nm and 330 nm, respectively. The excitation spectra of compound **5** was obtained by recording the emission at wavelength 450 nm. In all the cases, the excitation and emission slit widths were kept at 5 nm each. Compound **5** fluorescence emission or excitation spectra as a function of protein concentration were recorded by simple titration method (Banerjee et al., 2012; Banerji et al., 2013a, 2014; Maity et al., 2014; Ray et al., 2012).

Fluorescence anisotropy. Fluorescence anisotropy experiments were performed in the Cary Eclipse Fluorescence Spectrophotometer and a manual polarizer accessory was used. The excitation and emission wavelengths were set to 330 nm and 450 nm, respectively, with slit widths of 5nm for each monochromator. Anisotropy (r) was determined using the following equation (Banerjee et al., 2012):

$$r = (I_{VV} - G*I_{VH}) / (I_{VV} + 2G*I_{VH}); \; G = I_{HV}/I_{HH} \qquad \text{Eq. 4-1}$$

where I is the fluorescence emission intensity. The suffix V (vertical) or H (horizontal) denotes the alignment of excitation or emission polarizers. G is a correction factor. The changes in compound **5** fluorescence anisotropy as a function of protein concentration were recorded by titration method as mentioned earlier.

Determination of binding constants. K_d for compound **5** binding with BSA and HSA were determined from the compound **5** fluorescence anisotropy perturbation with proteins. Compound **5** concentration was kept at 0.5 μM and the protein concentration was varied from 0 μM to 5.5 μM. Small dilution error due to

the titration was ignored. Anisotropy values as a function of protein concentration were recorded. To derive the binding parameters, data were analyzed using the non-linear Langmuir isotherm (Banerji et al., 2013a):

$$\Delta r = \Delta r_{max} * [P] / (K_d + [P]) \quad \text{Eq. 4-2}$$

Where Δr is the difference in fluorescence anisotropy in the absence and presence of the protein at concentration [P], Δr_{max} is the maximum possible change in the fluorescence anisotropy, K_d is the binding dissociation constant. The non-linear equation was fitted to the data using Wolfram Mathematica 9.

Stern–Volmer quenching constant or the binding affinity constant, K_a was determined as a reciprocal of K_d (Banerji et al., 2013a).

Binding thermodynamics. K_d values were determined as a function of temperature and the thermodynamic parameters of binding was obtained by fitting van't Hoff equation (Banerji et al., 2013a; Ray et al., 2012) to the data:

$$\ln K_{eq} = -\Delta H^\circ/RT + \Delta S^\circ/R \quad \text{Eq. 4-3}$$

Where K_{eq} is the equilibrium constant (here the Stern–Volmer quenching constant) of binding at corresponding temperature T, and R is the gas constant. The equation gives the standard enthalpy change (ΔH°) and standard entropy change (ΔS°) on binding. The free energy change (ΔG°) has been estimated from the following relationship (Banerji et al., 2013a; Ray et al., 2012):

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \quad \text{Eq. 4-4}$$

Thermal and chemical denaturation. Thermal denaturation of protein was performed by increasing the temperature of the compound **5**-protein solution from 10 °C to 80 °C in 8 steps (Banerjee et al., 2012). The sample was kept under continuous stirring condition and the changes in compound **5** fluorescence spectra

were recorded. Chemical denaturation was achieved by increasing the concentration of urea (Banerjee et al., 2012) while recording the compound **5** fluorescence.

Lipophilicity and solubility calculations. Lipophilicity in terms of calculated logP (clogP) and solubility in terms of calculated logS (clogS) were determined at Virtual Computational Chemistry Laboratory (Tetko et al., 2005) server (http://www.vcclab.org/lab/alogps). Polar surface area was calculated with a 1.4 Å radius probe size.

Molecular docking. Molecular docking experiments were performed using AutoDock 4.2 (Morris et al., 2009) and AutoDock Vina (Trott and Olson, 2010) of The Scripps Research Institute and the SwissDock server (http://www.swissdock.ch) (Grosdidier et al., 2011). AutoDock 4.2, Vina and SwissDock uses different algorithm and scoring function for docking calculations. AutoDockTools (Morris et al., 2009) was used for the preparation of ligands and proteins for docking. BSA (PDB: 3V03) and HSA (PDB: 1E78) structural information were obtained from Protein Data Bank (Berman et al., 2000). The ligand structures were drawn in Avogadro (Hanwell et al., 2012) and geometry optimized *in vaccuo* using the steepest descent followed by conjugate gradient algorithms. Genetic algorithm was used in AutoDock 4.2 and it was run (ga_run) 100 times to generate a statistically significant number of docked poses (Alam et al., 2012). AutoDock 4.2 results were clustered using binding free energy and standard deviation cut offs of 0.5 kcal mol^{-1} and 2Å, respectively.

RESULTS AND DISCUSSION

Absorbance and fluorescence of compound 5. In aqueous buffer at pH 7.4, Compound **5** shows a strong absorption band with a peak at 330 nm. Compound **5** is also fluorescence active and the fluorescence band appeared at ~450 nm as shown in figure 4-1. Figure 4-1 also shows the excitation spectrum of compound **5**, which largely overlapped with the absorption spectrum suggesting that excited

state conformation of compound **5** in solution is homogeneous and close to ground state structure.

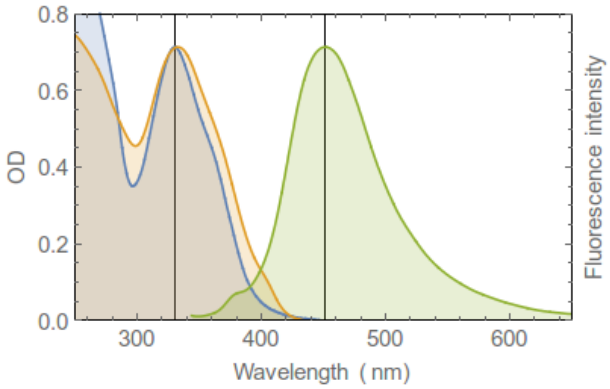


Figure 4-1: Absorption and fluorescence behavior of compound 5 in the UV-Visible range. Absorption spectrum of compound **5** in 20 mM Tris-HCl buffer of pH 7.4 is shown in blue. Fluorescence emission and excitation spectra of compound 5 are shown in green and yellow, respectively. Compound **5** fluorescence spectra were recorded in the same buffer and normalized against its absorption spectrum.

Fluorescence quantum yield of compound 5 in protein environment. Compound **5** absorbs UV-visible light strongly at the wavelengths where protein absorbs. It also has a strong absorption band in the fluorescence emission range of protein. Therefore, protein intrinsic fluorescence perturbation with compound **5** or the energy transfer from protein to compound **5** is not an eminent choice to probe the binding interactions of compound **5** with proteins. However, the changes in compound **5** fluorescence may be monitored as a parameter of binding. With the increasing concentration of serum albumins, we have found that the fluorescence intensity of the compound **5** increases (Figure 4-2A and Appendix II: Figure S4-1A). A blue shift in the emission maximum was also observed. It indicated compound **5** binding to the hydrophobic grooves of serum albumins. Such binding causes solvent exclusion of compound **5** and the energy that is otherwise spent in solvent relaxation, is gained by the emitting photons. Here, about 30 nm decrease in the Stokes' shift in compound **5** fluorescence corresponds to ~0.2 eV energy gain by each emitted photon.

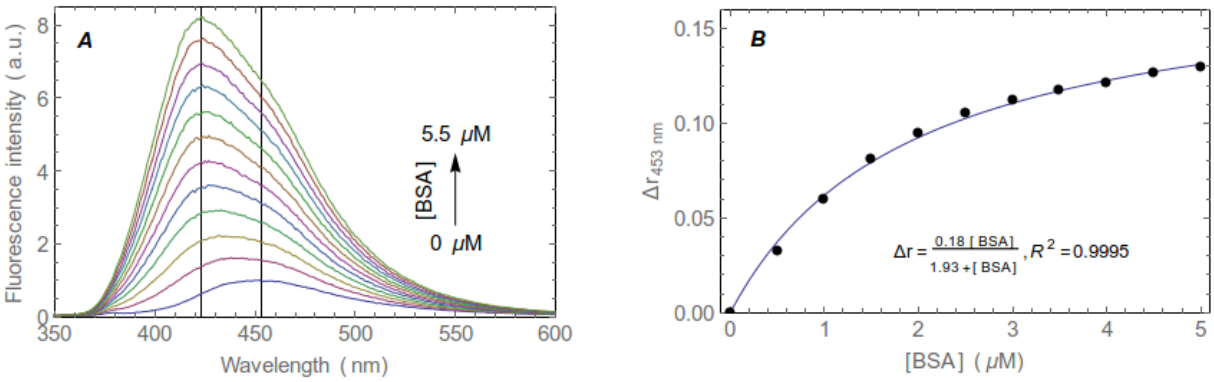


Figure 4-2: Fluorescence emission and anisotropy change of compound 5 in presence of serum albumin. Compound **5** concentration was kept constant at 0.5 μM and the protein concentration was varied from 0 through 5.5 μM. (A) The changes in fluorescence emission spectrum of compound **5** as a function of BSA concentration. (B) The changes in compound **5** fluorescence anisotropy with increasing concentration of BSA and the fitted Langmuir isotherm.

Interaction of compound **5** with serum albumins increases its fluorescence quantum yield and compound **5** gets a hydrophobic environments while bound to the serum albumins. Thus, the changes in compound **5** fluorescence in the presence of proteins carry the information not only about the interaction but the microenvironment of the binding site on its target protein as well.

Fluorescence anisotropy of compound 5 in presence of proteins. Small molecules tumble faster in less viscous solvents. But when it binds to a large molecule such as protein, its movement gets restricted. Fluorescence anisotropy is, therefore, widely used to measure the binding constants and kinetics of reactions that cause a change in the rotational time of the fluorescent molecules (Heyduk et al., 1996). Fluorescence anisotropy measurements can also elucidate the microenvironment of a small molecule in terms of its rotational diffusion, interactions, and proximity to proteins. Compound **5** in buffer solution shows very low anisotropy. With the increasing concentration of serum albumins, fluorescence anisotropy of compound **5** increases and gradually reaches the saturation (Figure 4-2B and Appendix II: Figure S4-1B). Binding dissociation constants for compound **5** binding with the two serum albumins were determined from this experiment and were found to be in the low micromolar concentration range (Table 4-1). The

molecule showed increased quantum yield along with a blue shift in presence of protein. Anisotropy experiment confirmed that the molecule goes inside the binding cavity of the protein, thus, restricting its free rotation. Therefore, the anisotropy suggests that the observed change in the fluorescence property of the molecule is a direct effect of the binding site environment.

Table 4-1: Binding constants of compound 5. The K_d and K_a values for the binding of the compound 5 to serum albumins as determined by the fluorescence and anisotropy perturbation experiments at room temperature.

Protein	K_d (M)	K_a (M^{-1})
BSA	1.93×10^{-6}	5.18×10^{5}
HSA	2.05×10^{-6}	4.88×10^{5}

Table 4-2: Binding thermodynamics. Thermodynamics of compound 5 binding to serum albumins.

Serum albumins	ΔG° (kJ mol^{-1}) at 25 °C	ΔH° (kJ mol^{-1})	ΔS° (J mol^{-1} K^{-1})
BSA	-31.64	21.47	178.25
HSA	-32.09	-24.28	26.18

Thermodynamics of compound 5 binding to serum albumins. Equilibrium constant of a reaction changes with the temperature (Figure 4-3). Such a change can be explained by van't Hoff's equation, which in turn, gives the standard enthalpy and standard entropy changes for the reaction. The associations of the compound with serum albumins are thermodynamically favorable, which is evident from the decrease in Gibbs free energy (Table 4-2). Moreover, the binding with HSA is enthalpy driven (negative ΔH°) whereas the binding with BSA is entropy driven (positive ΔS°). It suggests that, despite the structural similarity in the two proteins, the interactions with HSA are thermodynamically different from that of BSA.

Excited state geometry of compound 5 in protein environment. Fluorophore binding to a protein often results in an altered ground state electronic property, which can be visualized by a change in the absorption spectrum. However, the absorption spectrum of compound **5** does not change due to its binding with the serum albumins (Appendix II: Figure S4-2). It indicates that the ground state

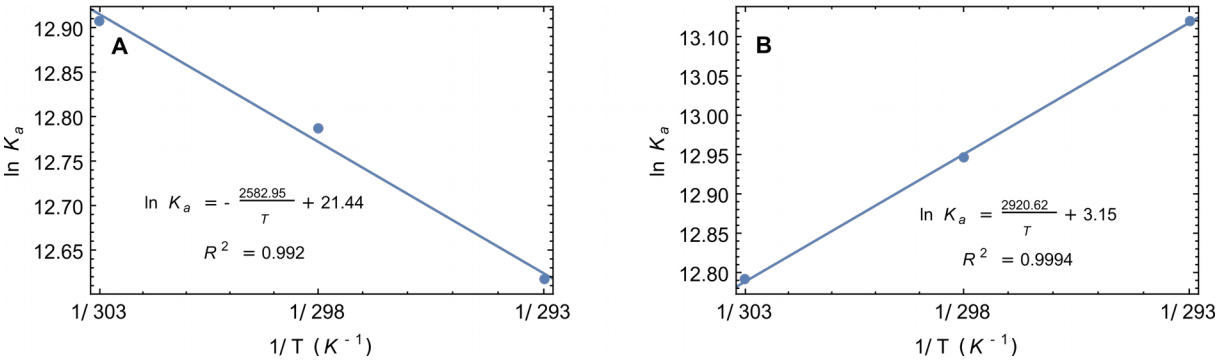


Figure 4-3: Determination of thermodynamic parameters of compound 5 binding from van't Hoff's plot. (A) Decrease in the binding equilibrium constant with the decreasing temperature for BSA-compound **5** interaction and the fitted van't Hoff equation. (B) Increase in the binding equilibrium constant with the decreasing temperature for HSA-compound **5** interaction and the fitted van't Hoff equation.

geometry of compound **5** inside the hydrophobic groove of serum albumins remains the same as in the solution. Interestingly, in the excitation spectrum of compound **5,** in presence of serum albumins (Figure 4-4 and Appedix II: Figure S4-3), ~8 nm red shift was observed (330 nm to 338 nm). It indicates that the excited state geometry of compound **5** within the binding site of serum albumin gets altered with the possible formation of an exciplex.

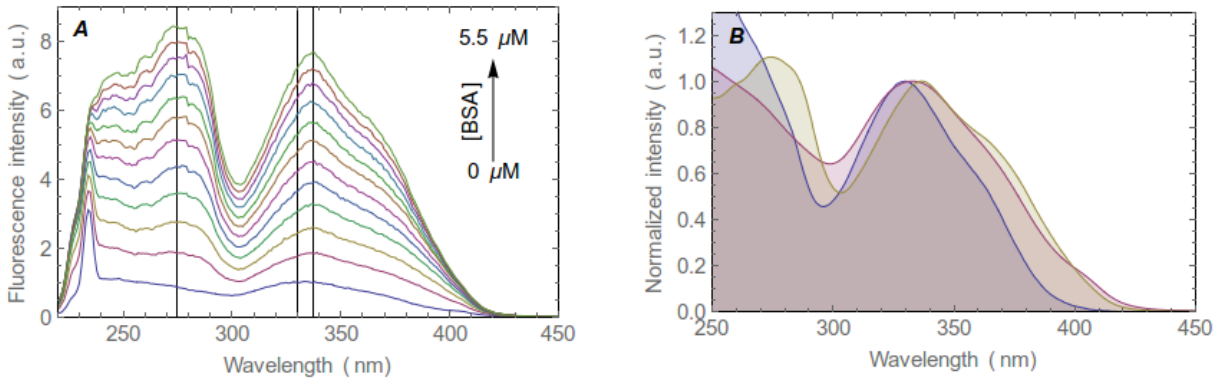


Figure 4-4: Compound 5 binding is an excited state phenomenon. (A) The changes in compound **5** fluorescence excitation spectra upon addition of BSA. (B) Normalized absorption (light blue) and fluorescence excitation (light red) spectra of compound **5** overlapped with fluorescence excitation spectrum (light yellow) in presence of BSA.

Denaturation of proteins bound to compound 5. Serum albumin-compound **5** complexes were denatured chemically and thermally. When

compound **5** is bound to the protein, fluorescence is blue shifted. With the gradual increase in the temperature or the urea concentration, the fluorescence intensity of compound **5** gradually decreases and the Stokes' shift increases until the fluorescence returns to its solution state nature (Figure 4-5 and Appendix II: Figure S4-4). This experiment demonstrated that compound **5** binds to the structured protein and not to a denatured protein. In other words, it reports the progressive loss of binding sites on its receptor when the receptor is undergoing a massive structural change.

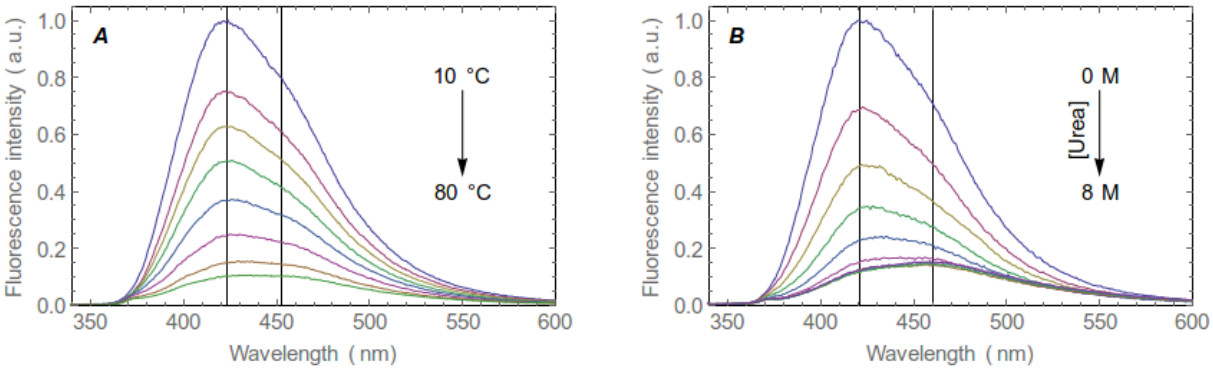


Figure 4-5: Compound 5 interaction with denaturing proteins. (A) The changes in fluorescence of BSA compound **5** complex with the increasing temperature. (B) The changes in fluorescence of BSA compound **5** complex with the increasing concentration of Urea.

Drug like properties of compound 5. The molecular properties of the compound, such as clogP, clogS, polar surface area etc. (Bickerton et al., 2012) are listed in the table 4-3. The clogP value of a compound is the logarithm of its partition coefficient between n-octanol and water. It is a well established measure of the compound's lipophilicity, which influences its behaviour in a range of biological processes such as solubility, membrane permeability, lack of selectivity and non-specific toxicity (Alam et al., 2011). It has been shown for compounds to have a reasonable probability of being well absorbed, their logP value must not be greater than 5.0 (Lipinski et al., 1997). Besides, the aqueous solubility of a compound is also defined by logS, which significantly affects its absorption and distribution characteristics. Typically, a low solubility goes along with a bad absorption. Most of the drugs on the market have an estimated logS value of about

-4. Table 3 lists the polar surface area of the compound as well, which should be less than 140 $Å^2$ for a drug molecule (Lipinski et al., 1997). Apart from lipophilicity/solubility and the polar surface area, the molecular weight and the number of hydrogen bond acceptor/donor in compound **5** also follow the Lipinski's rule of five (Lipinski et al., 1997) to be a candidate drug molecule.

Table 4-3: Molecular properties of compound 5.

Molecular properties	Values
logP#	4.56±0.62
logS	-5.97
Polar surface area	94.17 $Å^2$
Lipinski's rule of five	yes

The data represent mean ± SD.

Table 4-4: Theoretical binding free energies. Binding energies obtained by molecular docking experiments using three different algorithms, AutoDock 4.2, AutoDock Vina and SwissDock.

Protein	AutoDock 4.2 (kJ mol^{-1})‡	AutoDock Vina (kJ mol^{-1})	SwissDock (kJ mol^{-1})
BSA	-17.08±0.35	-33.05	-36.15
HSA	-20.78±0.44	-35.56	-35.52

‡ The data represent mean ± SEM.

Molecular modeling. *In silico* molecular docking calculation shows that the interactions of the compound with serum albumins are thermodynamically favorable (Table 4-4). The binding free energies computed by AutoDock Vina and SwissDock are very similar to that of the experimentally obtained values (Table 4-2). Molecular docking also provides the insight into the most favorable binding site for these compounds on the serum albumins. The lowest energy complexes obtained by the three different algorithms consistently showed that the binding sites for compound **5** lay in the groove between domain I and domain III of BSA, whereas it was within the domain I in case of HSA (Figure 4-6 and Appendix II: Figure S4-5). This may, in part, explain the enthalpy driven nature of binding with HSA and the entropy driven binding with BSA (Table 4-2). Moreover, the non-specific nature of the binding is apparent from the lack of clustering in the docking

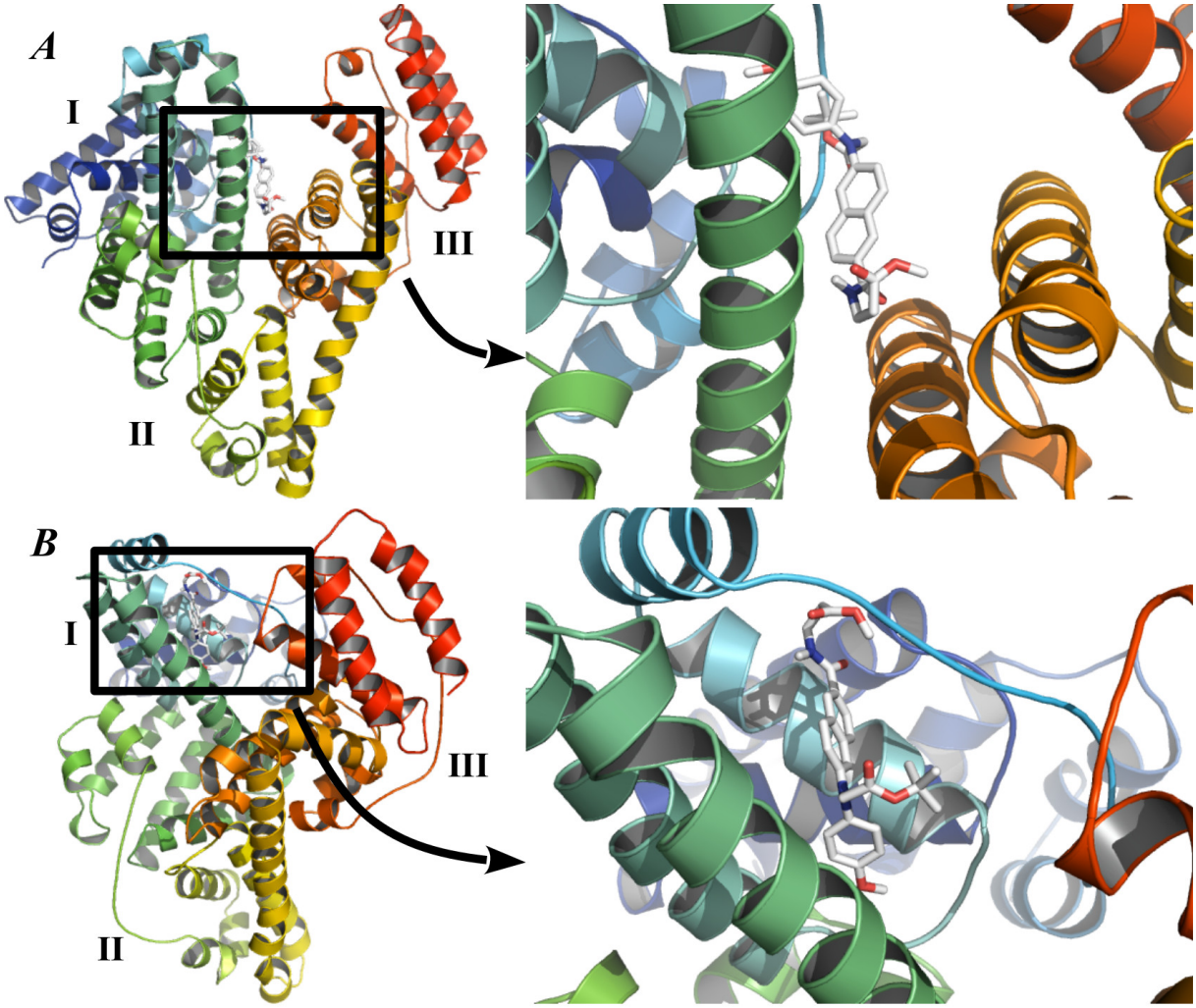


Figure 4-6: Interaction of compound 5 with serum albumins as obtained from molecular docking experiments. (A) Best binding conformation of compound **5** with BSA and the close up view. (B) Best binding conformations of compound **5** with HSA; it is also shown in close up. AutoDock Vina generated complexes are depicted here. Proteins are shown in ribbon diagram and the ligands in stick model. The three domains of serum albumin are marked with I—III. Standard color representation is used to denote the elements, H, N and O in the ligand.

results (Appendix II: Figure S4-6). We have shown in earlier works that the low energy high frequency clusters in docking output signifies specificity in the binding interactions (Alam et al., 2012; Bhowmik et al., 2013; Rudra et al., 2012). Serum albumin with its many hydrophobic binding pockets acts like a universal receptor for almost all drug molecules. Binding to serum albumin is generally non-specific in nature and driven by mainly hydrophobic interactions, which is evident in the molecular docking results as well (Appendix II: Figures S4-7 and S4-8).

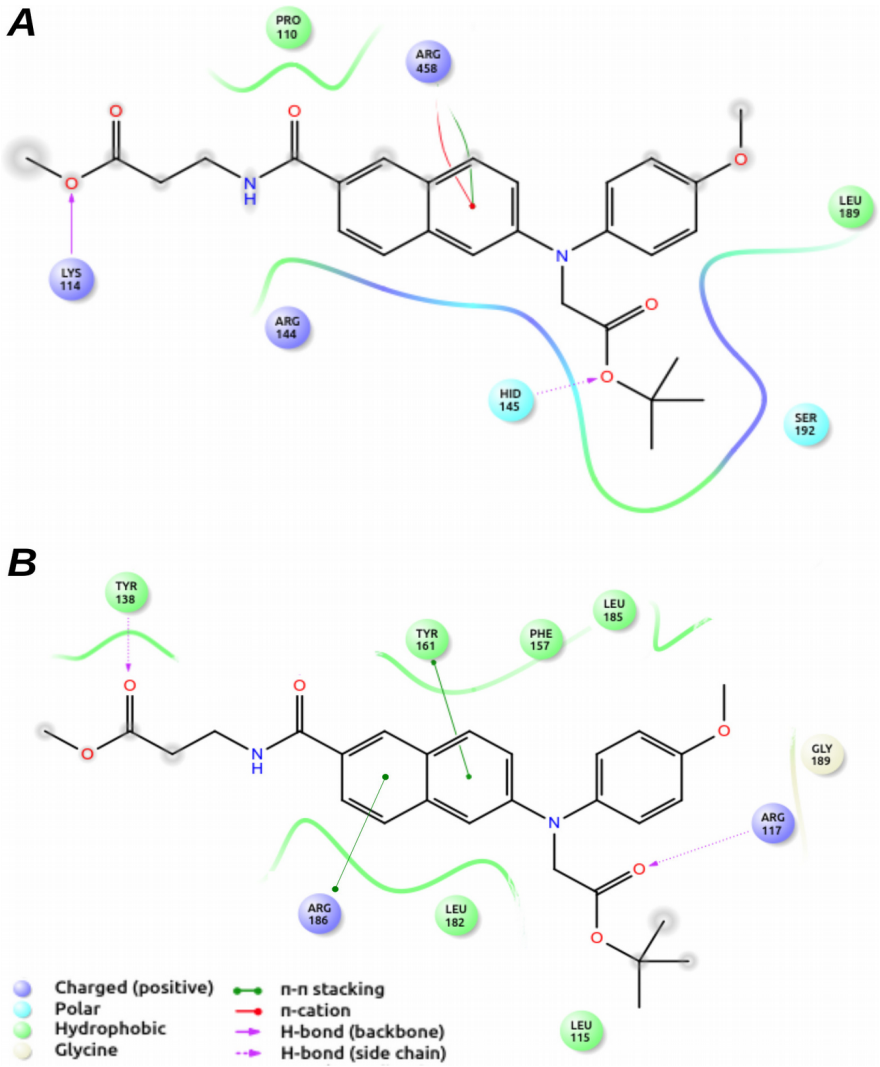


Figure 4-7: Detailed interaction diagram of compound 5 with serum albumins as obtained by molecular docking experiments. Only the residues common in all three lowest energy complexes generated by AutoDock 4.2, AutoDock Vina and SwissDock are shown. (A) Interacting residues of BSA and the types of interactions with compound **5**. (B) Interacting residues of HSA and the types of interactions with compound **5**.

Detailed interaction diagrams of the protein-ligand complexes showing the interacting residues and the types of interactions obtained by three different docking programs (AutoDock, Vina and SwissDock) are given the Appendix II (Figure S4-7 and S4-8). The consensus of the interacting residues of BSA and HSA with compound **5** are produced from those interacting diagrams depicted in figure S4-7 and S4-8 (*vide* Appendix II) and is shown in Figure 4-7. We have found that Lys114 and His145 of BSA forms H-Bond with compound 5, whereas, Arg458 forms

pi-cation interaction. Other important interacting residues of BSA are Arg144, Ser192, Pro110 and Leu189. In case of HSA, it is found that Arg117 and Tyr138 forms H-bond, whereas, Arg186 and Tyr161 forms pi-stacking with compound **5**. Other important interacting residues of HSA are Leu115, Phe157, Leu182, Leu185 and Gly189.

CONCLUSIONS

We have reported here, the spectroscopic behavior and binding parameters of a novel synthetic fluorophore in aqueous buffer and in the presence of albumin proteins. The compounds showed drug like properties and bind to serum albumin with the binding constants in low micromolar concentration range. Moreover, the compound showed some interesting properties that could be used to probe the microenvironment, which reflect the immediate surroundings of the molecule inside the target proteins. The compound binding to the hydrophobic sites of serum albumin significantly increased its fluorescence quantum yield, caused significant decrease in the Stokes' shift indicating changes of excited state geometry of the molecule inside protein binding pocket. We observed an overall effect of the microenvironoment of the binding site on the fluorophore. Whether it was hydrophobicity alone or some other factor such as polarity or charge or a combination of all and how those environmental variables correlate with the fluorescence property of the molecule requires further elaborate experiment to understand. In the future experiment we plan to study the fluorescence properties of this molecule in different solution condition, micellar and liposomal or membranous environments.

CHAPTER 5 | PROBING THE STRENGTH OF HYDROGEN BONDING WITH SOLUTION STATE NMR: A KEY TO THE LIGAND PROTEIN INTERACTIONS

Keywords: deuterium isotope effect, chemical shift, negative hyperconjugation, hydroxyl hydrogen bond, hexafluoroisopropanol.

INTRODUCTION

H-bonding involving hydroxyl (O-H) group has been recognized playing a fundamental role in the determination of structure, function, stability and dynamics of many chemical and biological systems (Jeffrey and Saenger, 1994; Perrin and Nielson, 1997). Intra- and inter-molecular H-bonds provide ample structural stability to proteins and nucleic acids. It also plays an important role in enzyme catalysis and molecular recognition processes (Fersht, 1977, 1987; Hibbert and Emsley, 1991). Although the existence and importance of these hydrogen bonds are beyond doubt, methods to determine the energetic contribution of individual hydrogen bond to binding and catalysis are not well established.

The strength and many other physical characteristics of conventional H-bonds involving O-H group are often measured following the structural and spectroscopic properties, such as chemical shift, coupling behavior etc. of the O-H group and the adjacent molecular components (Deshmukh and Gadre, 2009; Guo et al., 2012; Hankache et al., 2012; Hricovíni et al., 1997; Wendler et al., 2010). NMR and vibrational (FT-IR/Raman) spectroscopic methods has been, therefore, used extensively to quantify the parameters in nonaqueous solvents (D'Alva Torres et al., 1993; Rozenberg et al., 2003; Tonge et al., 1996). Direct characterization has been done by monitoring the O-H stretching frequency in the vibrational spectra of the H-bonded complex (Maiti et al., 2003). The presence of H-bond (involving O-H group as H donor) resulted red-shifts of the IR stretching frequencies of the O-H

group (Demmel et al., 1997). However, these characterizations are useful only for simple systems. Broad intense O-H stretching band from water hinders the measurement of specific O-H group of the macromolecules. To understand the hydrogen bonding interactions between enzyme and substrate selective alterations are made in the substrate/inhibitor or by site directed mutagenesis that eliminates a particular H-bond and the difference in binding affinity are often interpreted as the H-bond strength (Fersht, 1977, 1987, 1988; Fersht et al., 1985). However, this process fails to accommodate changes of solvent properties and the conformational effect.

Previous investigations by Raman spectroscopy and computational analysis have shown that O-H properties including its H-bonding characteristics and ionization behavior are reflected in the electronic properties of the associated C_α-H bond of the H-C_α-O-H functional group (Anderson, 2005; Gawlita et al., 2000; Jarmelo et al., 2005; Maiti et al., 2003, 2006). Quantum calculation by Gawlita et al. identified a correlation of C-H/D bond vibration frequency, $\nu_{C\text{-}H/D}$ at C2 of ethanol with H-bond formation (Gawlita et al., 2000). Experimentally it was verified and a significant red shift was found in $\nu_{C\text{-}H/D}$ due to H-bond formation of the H/D-C_α-O-H group of secondary alcohols (Maiti et al., 2003). A close analysis indicated that increase of H-bond strength correlated with a decrease of the H-C(OH) bond strength (Gawlita et al., 2000; Jarmelo et al., 2005; Maiti et al., 2003, 2006). It was further observed that $^1J_{CH}$ decreases ~0.2 Hz per kJ of H-bond strength (Maiti et al., 2006). Due to hydrogen bond formation delocalization of σ O-H bond occurred. It caused an increased overlap with the antibonding orbital of the C-H/D bond resulting in the reduction of the C_α-H/D bond order and weakening of C-H/D bond.

The average bond length is a manifestation of anharmonicity in the C-H/D stretching vibration. The weakening of the C-H or C-D bond of H/D-C_α-O-H due to involvement of the O-H group in H-bonding may be different due to possible differences in anharmonic factor encountered in C-H and C-D bonds. Deuterium substitution at the C–H hydrogen site leads to this isotope effects. This isotope effect is the manifestation of a small change in the vibrational state due to the

altered reduced mass upon deuteriation, and the changes in equilibrium geometry due to anharmonicity of the C–H stretching mode (Barfield and Fagerness, 1997; Chesnut and Foley, 1986; de Dios, 1996; Munch et al., 1992). One bond deuterium isotope effect on the ^{13}C chemical shift in a C-H system is defined as $^{1}\Delta^{13}C(D) = \sigma^{13}C(D)-\sigma^{13}C(H) = \delta^{13}C(H)-\delta^{13}C(D)$. In the current investigation we aimed to understand the anharmonic contribution to C_{α}-H group when O-H group in HFIP (1,1,1,3,3,3-hexafluoroisopropanol) is involved in H-bond donation via the measurement of deuterium isotope effect, changes in equilibrium distances and stretching frequencies. The calculated deuterium isotope effect on C2 of HFIP was compared with the experimentally observed values both in the monomeric state and when it involves in complex formation with tertiary amine. Due to moderate acidic nature of HFIP, it can form stronger H-bonds than many other alcohols. In addition, symmetry nature of the molecule minimizes the number of conformers in solution state. The presence of a single C-H bond and the previous success at determining a correlation between H-bond strength of different HFIP-tertiary amine complexes with $\nu_{C\text{-}H/D}$ and $^{1}J_{CH}$ in HFIP/HFIP-d2 added additional support to initiate the measurement of the deuterium isotope effect in free HFIP and when it forms complex with the amines (Maiti et al., 2003, 2006). Besides, the Raman and infrared spectra of HFIP and HFIP-d2 (2-deuterio-2-deuteriooxy-1,1,1,3,3,3-hexafluoropropane) have been previously characterized (Maiti et al., 2003; Murto et al., 1973). To calculate the isotope effect, nuclear shielding was calculated on DFT optimized structures using gauge independent atomic orbital (GIAO) method (Abildgaard et al., 1998; Lampert et al., 1997; Wolinski et al., 1990). The calculated isotope effect was compared to the experimental results. Both the calculation and experimental results showed a larger isotope effect in the amine complex and assigned as due to an increase in the anharmonicity of the weakened C-H/D bond.

MATERIALS AND METHODS

We carried out theoretical and experimental investigations on model systems of hydrogen bonded complexes between HFIP and tertiary amine. Deuterium isotope

effect on the chemical shift of secondary carbon of HFIP was monitored. Nuclear shielding were calculated on DFT optimized structures using GIAO method.

Chemicals. 1,1,1,3,3,3-hexafluuroisopropanaol (HFIP), 2-deuterio-2-deuteriooxy-1,1,1,3,3,3-hexafluoropropane (HFIP-d2) and other solvents were obtained commercially from Sigma Aldrich.

NMR spectroscopy. Hydrogen (H) coupled spectra were acquired using a Bruker 600 MHz spectrometer tuned to 13C and operating at 150.861 MHz. Data were recorded using a 40000 Hz sweep width, 4.2 μs. acquisition time 1.3 s with a delay of 3 s and referenced to internal DSS 2,2-dimethylsilapentane-5-sulfonic acid (DSS). The isotope effect was measured by the chemical shift differences in the 13C NMR resonance for C2 of HFIP and HFIP-d2. HFIP and HFIP-d2 were mixed 1:1 ratio (0.5 M) mixed in chloroform and triethylamine (TEA) and the proton coupled ^{13}C NMR spectra were recorded at room temperature.

Theoretical calculations. Minimum-energy structures of HFIP and its complex with trimethylamine (TMA) were obtained using B3LYP (Becke, 3-parameter, Lee-Yang-Parr) density functional method with the triple-ζ basis set 6-311G+(2d,p) (Andersson and Uvdal, 2005). All optimizations were performed without symmetry restrictions. Normal coordinate calculations were carried out at the minimum-energy geometries. The vibrational frequencies and all DFT (density functional theory) frequencies remained unscaled. All calculations were performed using the GAUSSIAN 09 program.

The NMR chemical shifts were calculated with the same basis set using the GIAO method on the optimized structure. The first derivative of the chemical shift with respect to the C-H bond lengths ($d\delta^{13}C/dr_{C\text{-}H}$) was calculated by shortening the C-H bond by 0.01 Å and recalculating the chemical shift.

The amount of the C-H bond shortening due to deuterium isotope substitution to the C-H ($\Delta r_{C\text{-}H/D}$) was calculated by scanning the C-H bond in the bond

direction at the B3LYP level in ten increments of 0.01 Å around the equilibrium position. A Morse function was fitted to the points that were below three times the zero-point energy (ZPE). The C-H bond length perturbation was calculated from the analytical solution to the Morse oscillator (Cheeseman et al., 1996; Nieto and Simmons, 1979).

RESULTS AND DISCUSSION

Geometry optimization. The molecular geometries of HFIP were optimized by density functional theory calculations, using the B3LYP hybrid functional and a 6-311G+(2d, p) basis set as implemented in Gaussian 09 (Andersson and Uvdal, 2005). The important bond (C-H, O-H) lengths and the characteristic dihedral angles of optimized geometry of HFIP and their complexes are shown in table 5-1. The optimized structure of HFIP-TMA (trimethylamine) complexes are shown in figure 5-1. The optimized coordinates are given in Appendix III. The energy difference between anti and gauche conformers was ~1.12 kcal mol^{-1}. The difference in energy between the two conformations decreased when HFIP formed H-bonded complexes with amines. Calculated and experimental chemical shifts ($\delta^{13}C$) (ppm) of different carbon atoms of HFIP are presented in table 5-2. Vibrational frequency v (cm^{-1}) and reduced mass μ (amu) for C-H/D bond of HFIP as a monomer and as its hydrogen bonded complex with TMA (gas phase) are shown in table 5-3.

The C_α-H bond lengths were increased in the amine complexes for both the anti and gauche conformations. The calculated geometries show that dihedral angles, H-C_α-O-H remains effectively constant in anti conformation after complex formation with trimethylamine and the C-H bond length increased by ~0.003 Å. For the gauche conformer similar C-H bond increasing was observed with a much effective change in H-C_α-O-H dihedral angle indicating different geometrical alignment in the complex. The increase of the C-H bond lengths were also reflected in their stretching vibrations (Table 5-3). The increase in bond length upon H-bond complex formation indicated weakening of C-H bond and rearrangement of the

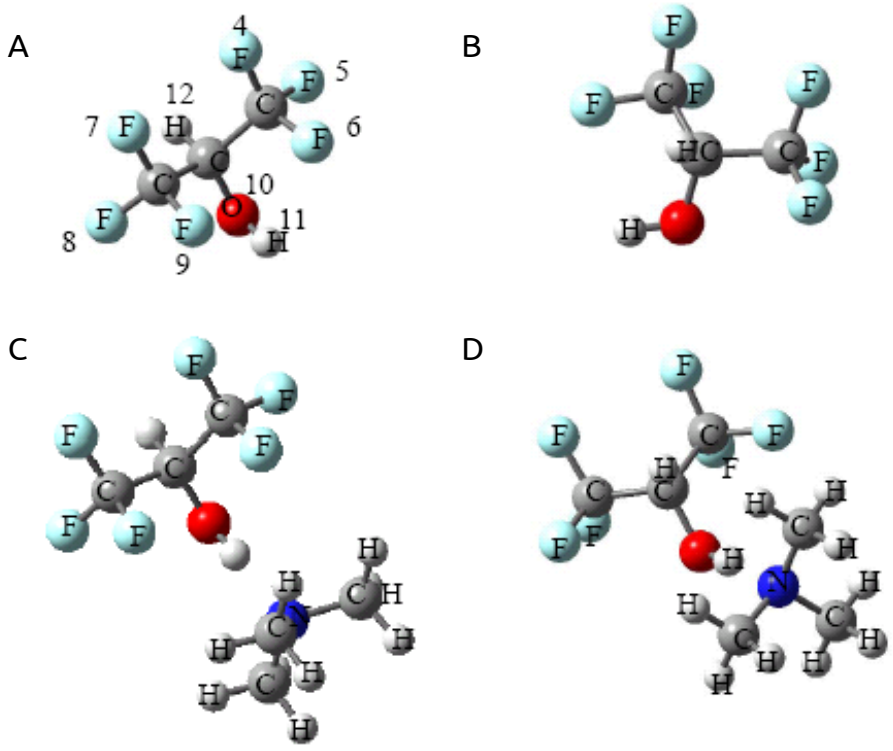


Figure 5-1: Ball and stick models of the energy minimized structure of HFIP and its H-bonded complexes with trimethylamine. (A) anti-HFIP, (B) gauche-HFIP, (C) anti-HFIP-TMA (D) gauche-HFIP-TMA. Atoms are marked and position are indicated in anti-HFIP (A).

electron cloud around the H-nucleus. The stretching vibration and the electron density around the nucleus are the major factor in the magnitude of the nuclear shielding.

13C Chemical Shifts and Its Derivatives. The chemical shift is calculated by applying the GIAO B3LYP method on optimized geometries. The calculated ^{13}C chemical shifts ($\delta^{13}C$) for the three carbon atoms of HFIP are given in Table 5-2. Two of the terminal carbon atoms (C1 and C3) of HFIP are of similar chemical environment and C2 is attached with both C-H and O-H bond and the chemical environment is quite different for this carbon atom. The calculated chemical shifts for C1 and C3 are identical for both the conformers. C2 shows slightly higher shielding values for both the anti and gauche conformations. The chemical shifts for other atoms are also included in the table. However, further calculation and experiments focused on C2 only.

Calculated chemical shifts for C1 and C3 were very close to each other for both the two conformers. The values were slightly higher when it formed the complex with the amine. The calculated chemical shifts are close to the experimental results (Table 5-2) and provided the confidence in the nuclear shielding calculation.

Table 5-1: Optimized parameters. Calculated energies, important bond lengths and dihedral angles for the DFT optimised structure of HFIP (anti and gauche conformers) and their complexes with trimethylamine.

Conformation	Energy (a.u.)	ZPE (a.u.)	Bond length (Å)		H-Cα-O-H
			C-H	O-H	angle (°)
anti-HFIP	-789.84598027	0.061779	1.09375	0.96974	180.00000
gauche-HFIP	-789.84420217	0.061526	1.09838	0.96688	-52.93699
anti-HFIP-TMA	-964.35693760	0.183872	1.09642	1.01404	177.28870
gauche-HFIP-TMA	-964.35654949	0.183748	1.09960	1.00937	-23.56804

Table 5-2: Theoretical and experimental chemical shifts. GIAO B3LYP/6-311+G(2d,p) calculated and experimental chemical shifts ($\delta^{13}C$) (ppm) of different carbon atoms of HFIP.

Atoms	HFIP			HFIP-TMA complex		
	Computed		Observed	Computed		Observed†
	anti	gauche		anti	gauche	
C1	134.679	134.918	121.345	135.838	135.957	122.717
C2	76.5125	77.0845	69.632	77.0751	75.6196	69.107
C3	134.679	133.577	121.345	135.999	135.066	122.717
F4	97.575	96.346		96.385	96.068	
F5	109.083	105.649		106.329	106.791	
F6	92	90.59		92.939	92.545	
F7	97.575	95.859		96.674	96.422	
F8	109.083	107.675		105.985	108.293	
F9	92	95.352		92.978	94.004	
O10	28.8499	28.3504		32.0546	33.4956	
H11	1.8129	1.634	7.264	9.8827	9.8345	7.359
H12	4.3	4.53	4.397	4.2731	4.2989	4.232

†HFIP-TEA complex was used for experiment.

Table 5-3: Computed frequencies. Vibrational frequency ν (cm-1) and reduced mass μ (amu) for C-H/D bond of HFIP as a monomer and as its hydrogen bonded complex with TMA (gas phase).

	HFIP				HFIP-TMA complex			
	C-H		C-D		C-H		C-D	
	anti	gauche	anti	gauche	anti	gauche	anti	gauche
ν	3118.4	3053.8	2300.7	2248.3	3081.5	3039.7	2268.9	2234.5
μ	1.0867	1.0848	2.3434	2.3330	1.0848	1.0840	2.3362	2.3303

The first derivative of the chemical shift $d\sigma/dr_{C\text{-}H}$ was calculated using the optimized geometries and by varying the C-H bond 0.01 Å and recalculating the chemical shift. Figure 5-2 shows the plot of chemical shift of C2 of anti-HFIP vs. C2-H bond distance of the optimized geometry (*vide* Materials and Methods). The slope in the linear plot gives the derivatives of the chemical shift. The first derivative of chemical shift of the C2 was -48.935 and -50.734 for anti and gauche conformations, respectively. However, in the complex the anti and gauche conformers showed chemical shifts of -50.275 and -50.755 ppm $Å^{-1}$, respectively.

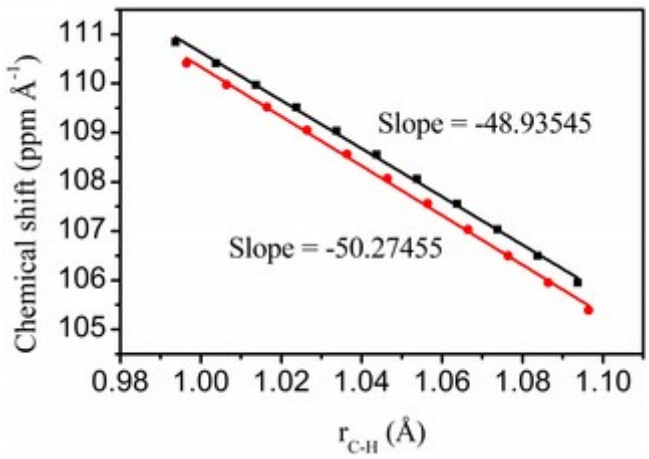


Figure 5-2: The first derivative of the chemical shift. Change in C2 nuclear shielding of HFIP in free (black) and complex form (red), upon C-H bond length perturbation.

Changes in the C-H bond length upon deuteriation. Due to isotope substitution at the C-H position, the average C-H/D bond length changed as a result of anharmonicity of the C-H/D stretching vibration. The results of normal mode

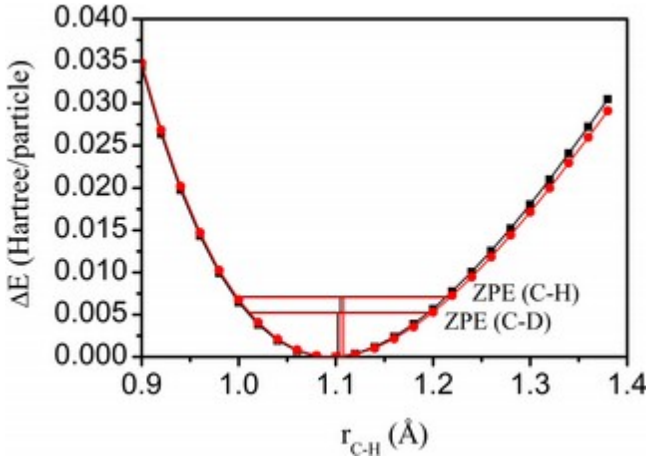


Figure 5-3: C-H/D bond length perturbation. Morse potential energy curve for C-H bond length perturbation with zero point energy (ZPE) correction in HFIP (black) and HFIP-TMA complex.

analysis, calculated reduced masses and vibrational frequencies are shown in table 5-3. The C-H stretching vibration for anti conformation was higher than gauche conformation by ~65 cm^{-1}. In the amine complex of HFIP the difference between the C-H vibration frequency decreased. The change in frequency between the two conformations was attributed to the geometrical alignment of the O-H and C-H groups. However, the interesting observation was that for any conformation in the complex form C-H stretching frequency was lower than the free monomer indicating enhanced anharmonicity in the C-H bond in the complex form.

The anhermonicity encountered in specific bond (specifically for diatomic molecule) could be analyzed by using Morse potential function represented by

$$V(r)=D_0[1-e^{-\alpha(r-r_e)}]^2 \qquad \text{Eq. 5-1}$$

Here r is the internuclear distance and re its equilibrium value. D_0 is the well depth, and α controls the width of the potential. The dissociation energy of the bond can be calculated by subtracting the zero point energy E_0 from the depth of the well. Energies are represented relative to the minimum; consequently all the energies are positive. A potential energy scan of the C-H bond stretching is described in the experimental section and shown in figure 5-3. By fitting the Morse function the values of D_0, α and r_e were obtained. The values of the parameters are

given in table 5-4. E_0 values for the anti-HFIP shows 0.00710 a.u. for C-H bond vibration in the ground state and reduced to 0.00524 a.u. for the deuterium substituted analogue. r_e was also decreased due to deuteriation. The difference in r_e due to deuterium substitution is given by Δr and the value was –0.00294 Å. Similar values were obtained for the gauche conformation.

Table 5-4: Isotope effect. Calculated values of isotope effect on C2 of HFIP in free and when it is associated with trimethylamine. C-H bond length 0.90 – 1.40 Å.

Parameters	anti		gauche	
HFIP/HFIP-d§				
r_0 (Å)	1.09375	1.09348	1.09838	1.09838
D_0 (a.u.)	0.17376	0.17357	0.15946	0.15946
α ($Å^{-1}$)	1.89204	1.89311	1.93492	1.93492
E_0 (a.u.)	0.00710	0.00524	0.00696	0.00512
r_e (Å)	1.10489	1.10195	1.11015	1.10706
Δr (Å)	-0.00294		-0.00309	
$d\delta^{13}C/dr_{C-H}$ (ppm $Å^{-1}$)	-48.935		-50.734	
$^1\Delta C(D)$ (ppb)	144		157	
HFIP-TMA complex/HFIP-d-TMA complex				
r_0 (Å)	1.09642	1.09641	1.09960	1.09960
D_0 (a.u.)	0.16498	0.16498	0.16498	0.15649
α ($Å^{-1}$)	1.91863	1.91863	1.94323	1.94323
E_0 (a.u.)	0.00702	0.00517	0.00693	0.00509
r_e (Å)	1.10800	1.10497	1.11144	1.10830
Δr (Å)	-0.00304		0.00314	
$d\delta^{13}C/dr_{C-H}$ (ppm $Å^{-1}$)	-50.275		-50.755	
$^1\Delta C(D)$ (ppb)	153		159	

§ HFIP-d, 2-deuterio-1,1,1,3,3,3-hexafluoroisopropanol

Figure 5-3 shows the calculated energy points for the C-H displacement in bond direction for the HFIP and HFIP-TMA complex. Morse potential function was fitted with the points, which are close to the zero point energy. All the fitted and calculated values are given in table 5-4. When HFIP forms complex with the amine equilibrium vibrational distance for C-H/D increases for both the conformations.

Compared to free monomer, in the complex formation deuterium substitution caused slightly higher r_e indicating enhanced anharmonicity.

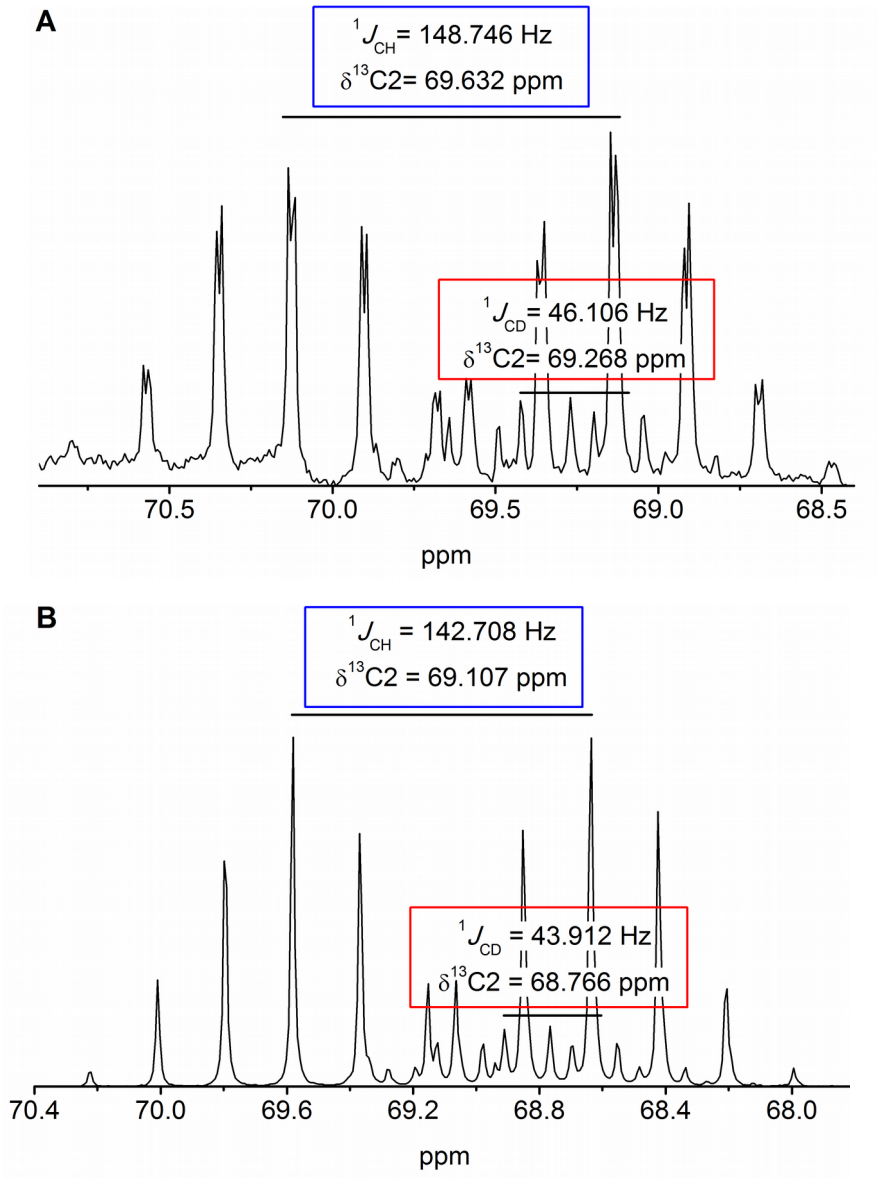


Figure 5-4: ^{13}C NMR spectra. (A) 600 MHz ^{13}C NMR spectra of mixture of HFIP and HFIP-d2 in chloroform. (B) 600 MHz ^{13}C NMR spectra of mixture of HFIP and HFIP-d2 in triethylamine. HFIP and HFIP-d2 was mixed at 1:1 ratio. The spectra are both fluorine and proton coupled. Only the C2 region is shown. Important peak positions are marked. One bond coupling behavior and isotope effects are also mentioned.

The isotope effect was calculated as the product of the first derivative of the chemical shift and the change in the bond length upon deuterium substitution

($d\delta^{13}C/dr_{C\text{-}H} \times \Delta r_{C\text{-}H/D}$). The values for the isotope effects are given in table 5-4. As shown in table 5-4 that the small changes in the r values causes change in the isotope effect. The C2 shielding gradient for the free monomer was ~1.3 unit less than when it formed the complex and the difference in r was ~0.003 Å. However, the small perturbations are important for isotope effect. The deuterium isotope effect increased in the complex, however, the magnitude was less. For the gauche conformation the calculated deuterium isotope effect on C2 was close to each other for both the monomer and when it formed the complex, However, there is a trend to higher isotope effect in complex formation.

NMR: chemical shift and isotope effect. Part of the ^{13}C NMR spectra of HFIP and HFIP-d2 mixtures are shown in Figure 5-4. Figure 5-4A is the $^{13}C2$ NMR spectra of HFIP and HFIP-d2 in chloroform. The spectra are both proton (and deuterium) and fluorine coupled. The intensity ratio and the splitting patterns for the signal for HFIP and HFIP-d2 were different due to difference in nuclear spin multiplicity of ^{1}H and ^{2}H.

C2 NMR signal becomes septet (seven splitting) due to coupling with six fluorine atoms attached to the adjacent carbon atoms (only the dominant five peaks are clearly visible in the figure 5-4). Further, coupling with H splits the septet into two groups of signals. Detailed analysis and the coupling pattern provided the chemical shift for C2 as 69.632 ppm. The coupling constant ($^{1}J_{CH}$) was 148.746 Hz. In HFIP-d2, septet of C2 signal was weaker due to substitution of deuterium and each bands in the septet split into three giving total 21 peak position. However only strong bands could be seen as shown in the figure. Close analysis of the C2 signal revealed that the chemical shift of C2 of HFIP-d2 was 69.286 and one bond $^{1}J_{CD}$ coupling was 46.106 Hz. The difference in the C2 chemical shift values was due to deuterium isotope effect. This small but certain change was possible to measure confidently by carrying out the experiment in a single cell and at similar condition.

Figure 5-4B shows the proton coupled C2 spectra of hydrogen bonded HFIP-amine complex. The spectrum patterns were similar to the monomers in

chloroform. Hydroxyl group (O-H/D) of HFIP/HFIP-d2 involved in H-bond donation to the tertiary N of triethylamine. In the amine solvent when it formed complex with the amine, δC2 = 69.107 (HFIP-TEA), δC2 = 68.766 ppm (HFIP-d2-TEA) and the difference in chemical shift, ΔδC2 was 341 ppb. The magnitude of the effect in the complex was slightly different compare to free monomer which was expected as the H-bonding induce more anharmonicity in the bonding pattern of the attached (either adjacent or further) nucleus.

A few experimental parameters exist that provide direct evidence of individual hydrogen bonds in complicated biological systems. Usually the existence of hydrogen bonds is inferred from the structure solved by either X-ray crystallography or NMR. However, mutation technology could help in deuterium substitution in C-H bond of any specific amino acid participating in enzymatic reaction (e.g., serine in protein kinases) and could help in measuring the isotope effect, thus, could be useful in determining the strength and nature of the H-bond. The anharmonicity encountered in the potential surface results in shorter average internuclear separation of the heavier isotopomer. A shorter bonds cause higher shielding, and hence a shift in the chemical shift. The chemical shifts for C2 of HFIP and HFIP-d2 were measured in the same experiment and mixing them in 1:1 ratio (Figure 5-4). In chloroform solvent the molecules remain as monomer. The isotope effect was 364 ppb. The gas phase calculation shows isotope effect of 144 ppb for anti conformation and 157 ppb for gauche conformation. The larger isotope effect observed in experimentally because of two reasons. The calculation done by substituting the single H in the C-H bond by deuterium, however in experiment C-D and O-D in HFIP-d2 replace both the C-H and O-H. Therefore, the larger isotope effect was anticipated in the measurement. Triethylamine was used in experiments whereas trimethylamine was used in calculations. In gas phase calculation, the solvation effect was ignored, which may also have immense influence on isotope effect (Buckingham et al., 2004; Homer, 1975).

In HFIP-amine complex hydroxyl group donates hydrogen and the O-H/D bond becomes weaker as it was reflected in the O-H stretching frequency. Not only

the O-H/D bond but also the C-H/D bond becomes weak and results in decrease of its stretching frequency. It has been recently established both by quantum mechanical calculation and experiments that the lone pair of O-H/D group may cause overlap with the C-H σ* (antibonding) orbital (known as negative hyperconjugation) resulting in weakening the C-H bond strength (Jarmelo et al., 2005; Maiti et al., 2003). However, the calculation showed higher isotope effect when HFIP form the H-bonded complex. Marshall pointed out that isotope effect might also occur due to the unequal motion of the centroid of the electron cloud of the bond and the nucleus (Marshall, 1961). For rapid vibration, the electron cloud may move differently and isotope effect may results. Therefore, the relative movement and the density of the electron cloud becomes important factors in isotope effect. The overlapping of the lone pair with of the C-H σ* orbital therefore also reflect in the deuterium isotope effect.

Theoretical study by Jameson and Osten established many useful relationships between the isotope effect and other bonding parameters (Jameson and Osten, 1986). Nuclear shielding depends on bond length and bond angles. The bond vibration and rotation lead to different average shift of their equilibrium positions. Therefore, the bond displacement becomes an important parameter in isotope effect.

The isotope effect may be caused directly by the effect of isotope on nuclear shielding or indirectly by the fact that substitution cause a change in the chemical equilibrium and consequently the nuclear shielding. Theoretically we observed that upon complex formation the r_e increased and the difference in vibrational equilibrium position also changed due to deuterium substitution. This was reflected in the H-bonded complex model in which O-H act as the proton donor to the amine bases and could account for the extra isotope effect on C2. The change in the shielding for different conformation and upon H-bond complex formation was not very dramatic with respect to absolute magnitude of the shielding constants. However, the change in the equilibrium position played an important role in the isotope effect.

CONCLUSIONS

To realize H-bonding property of a specific O-H group, the study aimed to capture and utilize unique anomeric effect (negative hyperconjugation) of the O-H group on the C_α-H/D bond (σ*). The deuterium substitution specifies the position. In addition to specificity it also provides a unique observable (physical parameter), the isotope effect which carries the information of H-bond properties and resultant anharmonicity. The dominant factor for the deuterium isotope effect was the change in the average vibrational distance upon deuteriation. The enhanced isotope effect in the H-bonded complex was due to enhanced anharmonicity which caused larger vibrational equilibrium distance. Change in vibrational equilibrium distance and the deuterium isotope effect could be used as the parameter in monitoring the strength of the H-bond in small model system with promising application in bio-macromolecules. The presence of a single hydrogen bond at an enzyme active site can be crucial to catalysis and the discussed parameters may be unique to determine the energetic contribution of individual hydrogen bonds to binding and catalysis.

"'I could have done it in a much more complicated
way,' said the red queen, immensely proud."
Lewis Carroll

REFERENCES

"It takes a thousand men to invent a telegraph,
or a steam engine, or a phonograph, or a photograph,
or a telephone or any other important thing..."
Mark Twain

Abildgaard, J., Bolvig, S., and Hansen, P.E. (1998). Unraveling the electronic and vibrational contributions to deuterium isotope effects on ^{13}C chemical shifts using ab initio model calculations. Analysis of the observed isotope effects on sterically perturbed intramolecular hydrogen-bonded o-hydroxy acyl aromatics. J. Am. Chem. Soc. 120, 9063–9069. DOI: 10.1021/ja9809051

Abou-Zied, O.K., and Al-Shihi, O.I.K. (2008). Characterization of subdomain IIA binding site of human serum albumin in its native, unfolded, and refolded states using small molecular probes. J. Am. Chem. Soc. 130, 10793–10801. DOI: 10.1021/ja8031289

Adessi, C., and Soto, C. (2002). Converting a peptide into a drug: strategies to improve stability and bioavailability. Curr. Med. Chem. 9, 963–978. DOI: 10.2174/0929867024606731

Ahmad, A., Uversky, V.N., Hong, D., and Fink, A.L. (2005). Early events in the fibrillation of monomeric insulin. J. Biol. Chem. 280, 42669–42675. DOI: 10.1074/jbc.M504298200

Alam, A., Pal, C., Goyal, M., Kundu, M.K., Kumar, R., Iqbal, M.S., Dey, S., Bindu, S., Sarkar, S., Pal, U., et al. (2011). Synthesis and bio-evaluation of human macrophage migration inhibitory factor inhibitor to develop anti-inflammatory agent. Bioorg. Med. Chem. 19, 7365–7373. DOI: 10.1016/j.bmc.2011.10.056

Alam, A., Haldar, S., Thulasiram, H.V., Kumar, R., Goyal, M., Iqbal, M.S., Pal, C., Dey, S., Bindu, S., Sarkar, S., et al. (2012). Novel anti-inflammatory activity of epoxyazadiradione against macrophage migration inhibitory factor: inhibition of tautomerase and proinflammatory activities of macrophage migration inhibitory factor. J. Biol. Chem. 287, 24844–24861. DOI: 10.1074/jbc.M112.341321

An, J., Totrov, M., and Abagyan, R. (2005). Pocketome via comprehensive identification and classification of ligand binding envelopes. Mol. Cell. Proteomics 4, 752–761. DOI: 10.1074/mcp.M400159-MCP200

Anderson, T.W., and Darling, D.A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann. Math. Stat. 23, 193–212. DOI: 10.1214/aoms/1177729437

Anderson, V.E. (2005). Quantifying energetic contributions to ground state destabilization. Arch. Biochem. Biophys. 433, 27–33. DOI: 10.1016/j.abb.2004.09.026

Andersson, M.P., and Uvdal, P. (2005). New scale factors for harmonic vibrational frequencies using the B3LYP density functional method with the triple-ζ basis set 6-311+G(d,p). J. Phys. Chem. A 109, 2937–2941. DOI: 10.1021/jp045733a

Apetri, M.M., Maiti, N.C., Zagorski, M.G., Carey, P.R., and Anderson, V.E. (2006). Secondary structure of α-synuclein oligomers: characterization by Raman and atomic force microscopy. J. Mol. Biol. 355, 63–71. DOI: 10.1016/j.jmb.2005.10.071

Arkin, M.R., and Wells, J.A. (2004). Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. Nat. Rev. Drug Discov. 3, 301–317. DOI: 10.1038/nrd1343

Ashburn, T.T., and Thor, K.B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. Nat. Rev. Drug Discov. 3, 673–683. DOI: 10.1038/nrd1468

Au, J.S.K., Cho, W.C.S., Yip, T., and Law, S.C.K. (2008). Proteomic approach to biomarker discovery in cancer tissue from lung adenocarcinoma among nonsmoking chinese women in Hong Kong. Cancer Invest. 26, 128–135. DOI: 10.1080/07357900701788031

Babu, M.M., van der Lee, R., de Groot, N.S., and Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. Curr. Opin. Struct. Biol. 21, 432–440. DOI: 10.1016/j.sbi.2011.03.011

Banerjee, M., Pal, U., Subudhhi, A., Chakrabarti, A., and Basu, S. (2012). Interaction of merocyanine 540 with serum albumins: photophysical and binding studies. J. Photochem. Photobiol. B 108, 23–33. DOI: 10.1016/j.jphotobiol.2011.12.005

Banerji, B., Pramanik, S.K., Pal, U., and Maiti, N.C. (2012). Conformation and cytotoxicity of a tetrapeptide constellated with alternative D- and L-proline. RSC Adv. 2, 6744–6747. DOI: 10.1039/C2RA20616A

Banerji, B., Pramanik, S.K., Pal, U., and Maiti, N.C. (2013a). Potent anticancer activity of cystine-based dipeptides and their interaction with serum albumins. Chem. Cent. J. 7, 91. DOI: 10.1186/1752-153X-7-91

Banerji, B., Pramanik, S.K., Pal, U., and Maiti, N.C. (2013b). Dipeptide derived from benzylcystine forms unbranched nanotubes in aqueous solution. J. Nanostructure Chem. 3, 12. DOI: 10.1186/2193-8865-3-12

Banerji, B., Pramanik, S.K., Pal, U., and Maiti, N.C. (2014). Binding of hemoglobin to ultrafine carbon nanoparticles: a spectroscopic insight into a major health hazard. RSC Adv. 4,

22536–22541. DOI: 10.1039/C4RA02569E

Barfield, M., and Fagerness, P. (1997). Density functional theory/GIAO studies of the ^{13}C, ^{15}N, and ^{1}H NMR chemical shifts in aminopyrimidines and aminobenzenes: relationships to electron densities and amine group orientations. J. Am. Chem. Soc. 119, 8699–8711. DOI: 10.1021/ja970990x

Barker, A., Kettle, J.G., Nowak, T., and Pease, J.E. (2013). Expanding medicinal chemistry space. Drug Discov. Today 18, 298–304. DOI: 10.1016/j.drudis.2012.10.008

Bartlett, J.B., Dredge, K., and Dalgleish, A.G. (2004). The evolution of thalidomide and its IMiD derivatives as anticancer agents. Nat. Rev. Cancer 4, 314–322. DOI: 10.1038/nrc1323

Berg, J., Tymoczko, J., and Stryer, L. (2002a). Section 22.5, Acetyl coenzyme A carboxylase plays a key role in controlling fatty acid metabolism. In Biochemistry, (New York: W.H. Freeman). http://www.ncbi.nlm.nih.gov/books/NBK22381/

Berg, J., Tymoczko, J., and Stryer, L. (2002b). Chapter 10, Regulatory strategies: enzymes and hemoglobin. In Biochemistry, (New York: W.H. Freeman). http://www.ncbi.nlm.nih.gov/ books/NBK21158/

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. 28, 235–242. DOI: 10.1093/nar/28.1.235

Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K., and Seitelberger, F. (1973). Brain dopamine and the syndromes of Parkinson and Huntington: clinical, morphological and neurochemical correlations. J. Neurol. Sci. 20, 415–455. DOI: 10.1016/0022-510X(73)90175-5

Bhowmik, A., Das, N., Pal, U., Mandal, M., Bhattacharya, S., Sarkar, M., Jaisankar, P., Maiti, N.C., and Ghosh, M.K. (2013). 2,2′-diphenyl-3,3′-diindolylmethane: a potent compound induces apoptosis in breast cancer cells by inhibiting EGFR pathway. PloS One 8, e59798. DOI: 10.1371/journal.pone.0059798

Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S., and Hopkins, A.L. (2012). Quantifying the chemical beauty of drugs. Nat. Chem. 4, 90–98. DOI: 10.1038/nchem.1243

Bjellqvist, B., Basse, B., Olsen, E., and Celis, J.E. (1994). Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis 15, 529–539. DOI: 10.1002/elps.1150150171

Blundell, T.L., Jhoti, H., and Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. Nat. Rev. Drug Discov. 1, 45–54. DOI: 10.1038/nrd706

Bordelon, T., Montegudo, S.K., Pakhomova, S., Oldham, M.L., and Newcomer, M.E. (2004). A disorder to order transition accompanies catalysis in retinaldehyde dehydrogenase type II. J. Biol. Chem. 279, 43085–43091. DOI: 10.1074/jbc.M406139200

Broeders, M.A.W., Doevendans, P.A., Bekkers, B.C.A.M., Bronsaer, R., Gorsel, E. van, Heemskerk, J.W.M., Egbrink, M.G.A. oude, Breda, E. van, Reneman, R.S., and Zee, R. van der (2000). Nebivolol: a third-generation β-blocker that augments vascular nitric oxide release endothelial β2-adrenergic receptor–mediated nitric oxide production. Circulation 102, 677–684. DOI: 10.1161/01.CIR.102.6.677

Brown, N., and Lewis, R.A. (2006). Exploiting QSAR methods in lead optimization. Curr. Opin. Drug Discov. Devel. 9, 419–424. PMID: 16889226

Buckingham, A.D., Schaefer, T., and Schneider, W.G. (2004). Solvent effects in nuclear magnetic resonance spectra. J. Chem. Phys. 32, 1227–1233. DOI: 10.1063/1.1730879

Burford, N.T., Watson, J., Bertekap, R., and Alt, A. (2011). Strategies for the identification of allosteric modulators of G-protein-coupled receptors. Biochem. Pharmacol. 81, 691–702. DOI: 10.1016/j.bcp.2010.12.012

Busto, E., Martínez-Montero, L., Gotor, V., and Gotor-Fernández, V. (2013). Chemoenzymatic asymmetric synthesis of serotonin receptor agonist (R)-frovatriptan. Eur. J. Org. Chem. 2013, 4057–4064. DOI: 10.1002/ejoc.201300114

Cardona, G., Rosselló, F., and Valiente, G. (2008). Extended Newick: it is time for a standard representation of phylogenetic networks. BMC Bioinformatics 9, 532. DOI: 10.1186/1471-2105-9-532

Cheeseman, J.R., Trucks, G.W., Keith, T.A., and Frisch, M.J. (1996). A comparison of models for calculating nuclear magnetic resonance shielding tensors. J. Chem. Phys. 104, 5497–5509. DOI: 10.1063/1.471789

Chen, J.J., Swope, D.M., Dashtipour, K., and Lyons, K.E. (2009). Transdermal rotigotine: a clinically innovative dopamine-receptor agonist for the management of Parkinson's disease. Pharmacotherapy 29, 1452–1467. DOI: 10.1592/phco.29.12.1452

Chen, J.W., Romero, P., Uversky, V.N., and Dunker, A.K. (2006). Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. J. Proteome Res. 5, 888–898. DOI: 10.1021/pr060049p

Cheng, B., Gong, H., Xiao, H., Petersen, R.B., Zheng, L., and Huang, K. (2013). Inhibiting toxic aggregation of amyloidogenic proteins: A therapeutic strategy for protein misfolding diseases. Biochim. Biophys. Acta BBA - Gen. Subj. 1830, 4860–4871. DOI: 10.1016/j.bbagen.2013.06.029

Cheng, Y., LeGall, T., Oldfield, C.J., Mueller, J.P., Van, Y.-Y.J., Romero, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2006). Rational drug design via intrinsically disordered protein. Trends Biotechnol. 24, 435–442. DOI: 10.1016/j.tibtech.2006.07.005

Cheng, Y., Oldfield, C.J., Meng, J., Romero, P., Uversky, V.N., and Dunker, A.K. (2007). Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochemistry 46, 13468–13477. DOI: 10.1021/bi7012273

Chesnut, D.B., and Foley, C.K. (1986). Chemical shifts and bond modification effects for some small first-row-atom molecules. J. Chem. Phys. 84, 852–861. DOI: 10.1063/1.450529

Christopoulos, A. (2002). Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. Nat. Rev. Drug Discov. 1, 198–210. DOI: 10.1038/nrd746

Clark, D.E., and Westhead, D.R. (1996). Evolutionary algorithms in computer-aided molecular design. J. Comput.

Aided Mol. Des. 10, 337–358. DOI: 10.1007/BF00124503

Clozel, M. (1991). Mechanism of action of angiotensin converting enzyme inhibitors on endothelial function in hypertension. Hypertension 18, II37. DOI: 10.1161/01.HYP.18.4_Suppl.II37

Cohen, B.E., McAnaney, T.B., Park, E.S., Jan, Y.N., Boxer, S.G., and Jan, L.Y. (2002). Probing protein electrostatics with a synthetic fluorescent amino acid. Science 296, 1700–1703. DOI: 10.1126/science.1069346

Cohlberg, J.A., Li, J., Uversky, V.N., and Fink, A.L. (2002). Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from alpha-synuclein in vitro. Biochemistry 41, 1502–1511. DOI: 10.1021/bi011711s

Cruz-Monteagudo, M., Medina-Franco, J.L., Pérez-Castillo, Y., Nicolotti, O., Cordeiro, M.N.D.S., and Borges, F. (2014). Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? Drug Discov. Today 19, 1069–1080. DOI: 10.1016/j.drudis.2014.02.003

Csizmók, V., Bokor, M., Bánki, P., Klement, E., Medzihradszky, K.F., Friedrich, P., Tompa, K., and Tompa, P. (2005). Primary contact sites in intrinsically unstructured proteins: the case of calpastatin and microtubule-associated protein 2. Biochemistry 44, 3955–3964. DOI: 10.1021/bi047817f

Cumberworth, A., Lamour, G., Babu, M.M., and Gsponer, J. (2013). Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. Biochem. J. 454, 361–369. DOI: 10.1042/BJ20130545

Curry, S., Mandelkow, H., Brick, P., and Franks, N. (1998). Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. Nat. Struct. Biol. 5, 827–835. DOI: 10.1038/1869

Dall'Acqua, S. (2014). Natural products as antimitotic agents. Curr. Top. Med. Chem. 14, 2272–2285. DOI: 10.2174/1568026614666141130095311

D'Alva Torres, G.S.F., Pouchan, C., Teixeira-Dias, J.J.C., and Fausto, R. (1993). Hydrogen bonding between substituted phenols and CH_3COOCH_3 or $CH_2ClCOOCH_3$: an FTIR spectroscopic study. Spectrosc. Lett. 26, 913–922. DOI: 10.1080/00387019308011581

Dalvit, C., Flocco, M., Knapp, S., Mostardini, M., Perego, R., Stockman, B.J., Veronesi, M., and Varasi, M. (2002). High-throughput NMR-based screening with competition binding experiments. J. Am. Chem. Soc. 124, 7702–7709. DOI: 10.1021/ja020174b

Das, S., Pal, U., Das, S., and Maiti, N.C. (2013). Chaperone action of cyclophilin on lysozyme and its aggregate. J. Proteins Proteomics 4, 129.

Demmel, F., Doster, W., Petry, W., and Schulte, A. (1997). Vibrational frequency shifts as a probe of hydrogen bonds: thermal expansion and glass transition of myoglobin in mixed solvents. Eur. Biophys. J. EBJ 26, 327–335. DOI: 10.1007/s002490050087

Dancheck, B., Nairn, A.C., and Peti, W. (2008). Detailed structural characterization of unbound protein phosphatase 1 inhibitors. Biochemistry 47, 12346–12356. DOI: 10.1021/bi801308y

Das, K., Martinez, S.E., Bandwar, R.P., and Arnold, E. (2014). Structures of HIV-1 RT-RNA/DNA ternary complexes with dATP and nevirapine reveal conformational flexibility of RNA/DNA: insights into requirements for RNase H cleavage. Nucleic Acids Res. 42, 8125–8137. DOI: 10.1093/nar/gku487

Das, S., Pal, U., Das, S., Bagga, K., Roy, A., Mrigwani, A., and Maiti, N.C. (2014). Sequence complexity of amyloidogenic regions in intrinsically disordered human proteins. PLoS ONE 9, e89781. DOI: 10.1371/journal.pone.0089781

De Clercq, E. (2002). Strategies in the design of antiviral drugs. Nat. Rev. Drug Discov. 1, 13–25. DOI: 10.1038/nrd703

De Smet, F., Christopoulos, A., and Carmeliet, P. (2014). Allosteric targeting of receptor tyrosine kinases. Nat. Biotechnol. 32, 1113–1120. DOI: 10.1038/nbt.3028

Deeks SG, Smith M, Holodniy M, and Kahn JO (1997). HIV-1 protease inhibitors: a review for clinicians. JAMA 277, 145–153. DOI: 10.1001/jama.1997.03540260059037

Deshmukh, M.M., and Gadre, S.R. (2009). Estimation of N–H···O═C intramolecular hydrogen bond energy in polypeptides. J. Phys. Chem. A 113, 7927–7932. DOI: 10.1021/jp9031207

van Dijk, A.D.J., Boelens, R., and Bonvin, A.M.J.J. (2005). Data-driven docking for the study of biomolecular complexes. FEBS J. 272, 293–312. DOI: 10.1111/j.1742-4658.2004.04473.x

de Dios, A.C. (1996). Ab initio calculations of the NMR chemical shift. Prog. Nucl. Magn. Reson. Spectrosc. 29, 229–278. DOI: 10.1016/S0079-6565(96)01029-1

Disfani, F.M., Hsu, W.-L., Mizianty, M.J., Oldfield, C.J., Xue, B., Dunker, A.K., Uversky, V.N., and Kurgan, L. (2012). MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. Bioinformatics 28, i75–i83. DOI: 10.1093/bioinformatics/bts209

Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21, 3433–3434. DOI: 10.1093/bioinformatics/bti541

Dosztányi, Z., Mészáros, B., and Simon, I. (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25, 2745–2746. DOI: 10.1093/bioinformatics/btp518

Dosztányi, Z., Mészáros, B., and Simon, I. (2010). Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. Brief. Bioinform. 11, 225–243. DOI: 10.1093/bib/bbp061

Dreyer, G.B., Metcalf, B.W., Tomaszek, T.A., Carr, T.J., Chandler, A.C., Hyland, L., Fakhoury, S.A., Magaard, V.W., Moore, M.L., and Strickler, J.E. (1989). Inhibition of human immunodeficiency virus 1 protease in vitro: rational design of substrate analogue inhibitors. Proc. Natl. Acad. Sci. 86, 9752–9756. PMID: 2690072

Duggan, P.J., and Tuck, K.L. (2015). Bioactive Mimetics of Conotoxins and other Venom Peptides. Toxins 7, 4175–4198. DOI: 10.3390/toxins7104175

Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., and

Brown, C.J. (2000). Intrinsic protein disorder in complete genomes. Genome Inform. 11, 161–171. PMID: 11700597

Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M., and Obradović, Z. (2002). Intrinsic disorder and protein function. Biochemistry 41, 6573–6582. DOI: 10.1021/bi012159+

Dunker, A.K., Cortese, M.S., Romero, P., Iakoucheva, L.M., and Uversky, V.N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. 272, 5129–5148. DOI: 10.1111/j.1742-4658.2005.04948.x

Durbin, R. (1998). Biological sequence analysis: probabilistic models of proteins and nucleic acids (Cambridge UK: Cambridge university press). ISBN 13: 9780521629713

Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr. Opin. Struct. Biol. 12, 54–60. DOI: 10.1016/S0959-440X(02)00289-0

Dyson, H.J., and Wright, P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197–208. DOI: 10.1038/nrm1589

Edfeldt, F.N.B., Folmer, R.H.A., and Breeze, A.L. (2011). Fragment screening to predict druggability (ligandability) and lead discovery success. Drug Discov. Today 16, 284–287. DOI: 10.1016/j.drudis.2011.02.002

Edwards, Y.J., Lobley, A.E., Pentony, M.M., and Jones, D.T. (2009). Insights into the regulation of intrinsically disordered proteins in the human proteome by analyzing sequence and gene expression data. Genome Biol. 10, R50. DOI: 10.1186/gb-2009-10-5-r50

Er, J.C., Vendrell, M., Tang, M.K., Zhai, D., and Chang, Y.-T. (2013). Fluorescent dye cocktail for multiplex drug-site mapping on human serum albumin. ACS Comb. Sci. 15, 452–457. DOI: 10.1021/co400060b

Fersht, A. (1977). Enzyme Structure and Mechanism (W.H. Freeman & Company). ISBN 13: 9780716701880

Fersht, A.R. (1987). The hydrogen bond in molecular recognition. Trends Biochem. Sci. 12, 301–304. DOI: 10.1016/0968-0004(87)90146-0

Fersht, A.R. (1988). Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. Biochemistry 27, 1577–1580. DOI: 10.1021/bi00405a027

Fersht, A.R., Shi, J.-P., Knill-Jones, J., Lowe, D.M., Wilkinson, A.J., Blow, D.M., Brick, P., Carter, P., Waye, M.M.Y., and Winter, G. (1985). Hydrogen bonding and biological specificity analysed by protein engineering. Nature 314, 235–238. DOI: 10.1038/314235a0

Foda, Z.H., and Seeliger, M.A. (2014). Kinase inhibitors: An allosteric add-on. Nat. Chem. Biol. 10, 796–797. DOI: 10.1038/nchembio.1630

Folmar, L.C., Hemmer, M., Hemmer, R., Bowman, C., Kroll, K., and Denslow, N.D. (2000). Comparative estrogenicity of estradiol, ethynyl estradiol and diethylstilbestrol in an in vivo, male sheepshead minnow (Cyprinodon variegatus), vitellogenin bioassay. Aquat. Toxicol. 49, 77–88. DOI: 10.1016/S0166-445X(99)00076-4

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47, 1739–1749. DOI: 10.1021/jm0306430

Frye, S.V. (2010). The art of the chemical probe. Nat. Chem. Biol. 6, 159–161. DOI: 10.1038/nchembio.296

Fukuchi, S., Sakamoto, S., Nobe, Y., Murakami, S.D., Amemiya, T., Hosoda, K., Koike, R., Hiroaki, H., and Ota, M. (2012). IDEAL: intrinsically disordered proteins with extensive annotations and literature. Nucleic Acids Res. 40, D507–D511. DOI: 10.1093/nar/gkr884

Fuxreiter, M., Simon, I., Friedrich, P., and Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol. 338, 1015–1026. DOI: 10.1016/j.jmb.2004.03.017

Fuxreiter, M., Tóth-Petróczy, Á., Kraut, D.A., Matouschek, A.T., Lim, R.Y.H., Xue, B., Kurgan, L., and Uversky, V.N. (2014). Disordered proteinaceous machines. Chem. Rev. 114, 6806–6843. DOI: 10.1021/cr4007329

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. 'everine, Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Humana Press), pp. 571–607. DOI: 10.1385/1-59259-890-0:571

Gawlita, E., Lantz, M., Paneth, P., Bell, A.F., Tonge, P.J., and Anderson, V.E. (2000). H-bonding in alcohols is reflected in the C_α–H bond strength: variation of C–D vibrational frequency and fractionation factor. J. Am. Chem. Soc. 122, 11660–11669. DOI: 10.1021/ja001891d

Goddard, J.-P., and Reymond, J.-L. (2004). Enzyme assays for high-throughput screening. Curr. Opin. Biotechnol. 15, 314–322. DOI: 10.1016/j.copbio.2004.06.008

Goodsell, D.S. (2014). Microtubules. RCSB Protein Data Bank. DOI: 10.2210/rcsb_pdb/ mom_2014_7

Goodsell, D.S. (2015). Amyloids. RCSB Protein Data Bank. DOI: 10.2210/rcsb_pdb/ mom_2015_9

Gray, D.M. (1996). Circular dichroism of protein-nucleic acid interactions. In Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, ed. (Springer US), pp. 469–500. DOI: 10.1007/978-1-4757-2508-7_13

Grosdidier, A., Zoete, V., and Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. Nucleic Acids Res. 39, W270–W277. DOI: 10.1093/nar/gkr366

Gsponer, J., Futschik, M.E., Teichmann, S.A., and Babu, M.M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science 322, 1365–1368. DOI: 10.1126/science.1163581

Gunasekaran, K., Tsai, C.-J., Kumar, S., Zanuy, D., and Nussinov, R. (2003). Extended disordered proteins: targeting function with less scaffold. Trends Biochem. Sci. 28, 81–85. DOI: 10.1016/S0968-0004(03)00003-3

Guo, J., Tolstoy, P.M., Koeppe, B., Golubev, N.S., Denisov, G.S., Smirnov, S.N., and Limbach, H.-H. (2012). Hydrogen bond geometries and proton tautomerism of homoconjugated

anions of carboxylic acids studied via H/D isotope effects on ^{13}C NMR chemical shifts. J. Phys. Chem. A 116, 11180–11188. DOI: 10.1021/jp304943h

Habchi, J., Tompa, P., Longhi, S., and Uversky, V.N. (2014). Introducing protein intrinsic disorder. Chem. Rev. 114, 6561–6588. DOI: 10.1021/cr400514h

Hajduk, P.J., Huth, J.R., and Fesik, S.W. (2005). Druggability indices for protein targets derived from NMR-based screening data. J. Med. Chem. 48, 2518–2525. DOI: 10.1021/jm049131r

Halper, J., and Kjaer, M. (2014). Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins and thrombospondins. Adv. Exp. Med. Biol. 802, 31–47. DOI: 10.1007/978-94-007-7893-1_3

Hankache, J., Hanss, D., and Wenger, O.S. (2012). Hydrogen-bond strengthening upon photoinduced electron transfer in ruthenium–anthraquinone dyads interacting with hexafluoroisopropanol or water. J. Phys. Chem. A 116, 3347–3358. DOI: 10.1021/jp300090n

Hantschel, O., Grebien, F., and Superti-Furga, G. (2012). The growing arsenal of ATP-competitive and allosteric inhibitors of BCR–ABL. Cancer Res. 72, 4890–4895. DOI: 10.1158/0008-5472.CAN-12-1276

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E., and Hutchison, G.R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J. Cheminformatics 4, 17. DOI: 10.1186/1758-2946-4-17

Hatano, M. (1986). In Induced Circular Dichroism in Biopolymer-Dye Systems (Springer Berlin Heidelberg). DOI: 10.1007/BFb0071111

Haynes, C., Oldfield, C.J., Ji, F., Klitgord, N., Cusick, M.E., Radivojac, P., Uversky, V.N., Vidal, M., and Iakoucheva, L.M. (2006). Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2, e100. DOI: 10.1371/journal.pcbi.0020100

Heinrich, M., and Lee Teoh, H. (2004). Galanthamine from snowdrop—the development of a modern drug against Alzheimer's disease from local Caucasian knowledge. J. Ethnopharmacol. 92, 147–162. DOI: 10.1016/j.jep.2004.02.012

Heyduk, T., Ma, Y., Tang, H., and Ebright, R.H. (1996). Fluorescence anisotropy: rapid, quantitative assay for protein-DNA and protein-protein interaction. In Methods in Enzymology, Sankar Adhya, ed. (Academic Press), pp. 492–503. DOI: 10.1016/S0076-6879(96)74039-9

Hibbert, F., and Emsley, J. (1991). Hydrogen bonding and chemical reactivity. In Advances in Physical Organic Chemistry, (London: Academic Press Limited), pp. 255–379. DOI: 10.1016/S0065-3160(08)60047-7

Homer, J. (1975). Solvent effects on nuclear magnetic resonance chemical shifts. Appl. Spectrosc. Rev. 9, 1–132. DOI: 10.1080/05704927508081488

Hopkins, A.L., and Groom, C.R. (2002). The druggable genome. Nat. Rev. Drug Discov. 1, 727–730. DOI: 10.1038/nrd892

Hricovíni, M., Malkina, O.L., Bízik, F., Nagy, L.T., and Malkin, V.G. (1997). Calculation of NMR chemical shifts and spin–spin coupling constants in the monosaccharide methyl-β-d-xylopyranoside using a density functional theory approach. J. Phys. Chem. A 101, 9756–9762. DOI: 10.1021/jp972071z

Hu, Y., and Bajorath, J. (2012). Extending the activity cliff concept: structural categorization of activity cliffs and systematic identification of different types of cliffs in the ChEMBL database. J. Chem. Inf. Model. 52, 1806–1811. DOI: 10.1021/ci300274c

Huang, K., Maiti, N.C., Phillips, N.B., Carey, P.R., and Weiss, M.A. (2006). Structure-specific effects of protein topology on cross-beta assembly: studies of insulin fibrillation. Biochemistry 45, 10278–10293. DOI: 10.1021/bi060879g

Idris, I., and Donnelly, R. (2009). Sodium-glucose co-transporter-2 inhibitors: an emerging new class of oral antidiabetic drug. Diabetes Obes. Metab. 11, 79–88. DOI: 10.1111/j.1463-1326.2008.00982.x

Ivetac, A., and Andrew McCammon, J. (2010). Mapping the Druggable Allosteric Space of G-Protein Coupled Receptors: a Fragment-Based Molecular Dynamics Approach. Chem. Biol. Drug Des. 76, 201–217.

Jackson, D.A., Symons, R.H., and Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc. Natl. Acad. Sci. 69, 2904–2909. PMID: 4342968

Jameson, C.J., and Osten, H.J. (1986). Theoretical aspects of isotope effects on nuclear shielding. In Annual Reports on NMR Spectroscopy, (London: Academic Press), pp. 1–75. DOI: 10.1016/S0066-4103(08)60234-3

Jarmelo, S., Maiti, N., Anderson, V., Carey, P.R., and Fausto, R. (2005). C_α–H bond-stretching frequency in alcohols as a probe of hydrogen-bonding strength: a combined vibrational spectroscopic and theoretical study of n-[1-D]propanol. J. Phys. Chem. A 109, 2069–2077. DOI: 10.1021/jp046683c

Jeffrey, G.A., and Saenger, W. (1994). Hydrogen Bonding in Biological Structures (Springer-Verlag). DOI: 10.1007/978-3-642-85135-3

Jiang, L., Gao, Y., Mao, F., Liu, Z., and Lai, L. (2002). Potential of mean force for protein-protein interaction studies. Proteins 46, 190–196. DOI: 10.1002/prot.10031

Johnson, M.A., and Maggiora, G.M. (1990). Concepts and Applications of Molecular Similarity (Wiley). ISBN 13: 9780471621751

Jolivet, J., Cowan, K.H., Curt, G.A., Clendeninn, N.J., and Chabner, B.A. (1983). The pharmacology and clinical use of methotrexate. N. Engl. J. Med. 309, 1094–1104. DOI: 10.1056/NEJM198311033091805

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202. DOI: 10.1006/jmbi.1999.3091

Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31, 857–863. DOI: 10.1093/bioinformatics/btu744

Jones, G., Willett, P., Glen, R.C., Leach, A.R., and Taylor, R. (1997). Development and validation of a genetic algorithm for

flexible docking. J. Mol. Biol. 267, 727–748. DOI: 10.1006/jmbi.1996.0897

Jung, Y., and Lippard, S.J. (2007). Direct cellular responses to platinum-induced DNA damage. Chem. Rev. 107, 1387–1407. DOI: 10.1021/cr068207j

Kabsch, W., and Vandekerckhove, J. (1992). Structure and function of actin. Annu. Rev. Biophys. Biomol. Struct. 21, 49–76. DOI: 10.1146/annurev.bb.21.060192.000405

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181, 662–666. DOI: 10.1038/181662a0

Kielty, C.M., Sherratt, M.J., and Shuttleworth, C.A. (2002). Elastic fibres. J. Cell Sci. 115, 2817–2828. PMID: 12082143

Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., et al. (2016). PubChem substance and compound databases. Nucleic Acids Res. 44, D1202–D1213. DOI: 10.1093/nar/gkv951

Kiraga, J., Mackiewicz, P., Mackiewicz, D., Kowalczuk, M., Biecek, P., Polak, N., Smolarczyk, K., Dudek, M.R., and Cebrat, S. (2007). The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. BMC Genomics 8, 163. DOI: 10.1186/1471-2164-8-163

Koehn, F.E., and Carter, G.T. (2005). The evolving role of natural products in drug discovery. Nat. Rev. Drug Discov. 4, 206–220. DOI: 10.1038/nrd1657

Kogej, T., Blomberg, N., Greasley, P.J., Mundt, S., Vainio, M.J., Schamberger, J., Schmidt, G., and Hüser, J. (2013). Big pharma screening collections: more of the same or unique libraries? The AstraZeneca–Bayer Pharma AG case. Drug Discov. Today 18, 1014–1024. DOI: 10.1016/j.drudis.2012.10.011

Kollár, P., Rajchard, J., Balounová, Z., and Pazourek, J. (2014). Marine natural products: bryostatins in preclinical and clinical studies. Pharm. Biol. 52, 237–242. DOI: 10.3109/13880209.2013.804100

Kötting, C., and Gerwert, K. (2013). Monitoring protein–ligand interactions by time-resolved FTIR difference spectroscopy. In Protein-Ligand Interactions, M.A. Williams, and T. Daviter, eds. (Humana Press), pp. 299–323. DOI: 10.1007/978-1-62703-398-5_11

Kramer, B., Rarey, M., and Lengauer, T. (1999). Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. Proteins Struct. Funct. Bioinforma. 37, 228–241. DOI: 10.1002/(SICI)1097-0134(19991101)37:2<228::AID-PROT8>3.0.CO;2-8

Krebs, M.R.H., Bromley, E.H.C., and Donald, A.M. (2005). The binding of thioflavin-T to amyloid fibrils: localisation and implications. J. Struct. Biol. 149, 30–37. DOI: 10.1016/j.jsb.2004.08.002

Krishnan, N., Koveal, D., Miller, D.H., Xue, B., Akshinthala, S.D., Kragelj, J., Jensen, M.R., Gauss, C.-M., Page, R., Blackledge, M., et al. (2014). Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. Nat. Chem. Biol. 10, 558–566. DOI: 10.1038/nchembio.1528

Kuffler, S.W., and Edwards, C. (1958). Mechanism of gamma aminobutyric acid (GABA) action and its relation to synaptic inhibition. J. Neurophysiol. 21, 589–610. PMID: 13599049

Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E. (1982). A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161, 269–288. DOI: 10.1016/0022-2836(82)90153-X

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132. DOI: 10.1016/0022-2836(82)90515-0

Lakowicz, J.R. (2006a). Quenching of fluorescence. In Principles of Fluorescence Spectroscopy, (Springer US), pp. 277–330. DOI: 10.1007/978-0-387-46312-4_8

Lakowicz, J.R. (2006b). Fluorescence anisotropy. In Principles of Fluorescence Spectroscopy, (Springer US), pp. 353–382. DOI: 10.1007/978-0-387-46312-4_10

Lakowicz, J.R. (2006c). Energy transfer. In Principles of Fluorescence Spectroscopy, (Springer US), pp. 443–475. DOI: 10.1007/978-0-387-46312-4_13

Lakowicz, J.R. (2006d). Time-resolved protein fluorescence. In Principles of Fluorescence Spectroscopy, (Springer US), pp. 577–606. DOI: 10.1007/978-0-387-46312-4_17

Lamba, V., and Ghosh, I. (2012). New directions in targeting protein kinases: focusing upon true allosteric and bivalent inhibitors. Curr. Pharm. Des. 18, 2936–2945. DOI: 10.2174/138161212800672813

Lampert, H., Mikenda, W., Karpfen, A., and Kählig, H. (1997). NMR shieldings in benzoyl and 2-hydroxybenzoyl compounds. Experimental versus GIAO calculated data. J. Phys. Chem. A 101, 9610–9617. DOI: 10.1021/jp970280d

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921. DOI: 10.1038/35057062

Lao, B.B., Drew, K., Guarracino, D.A., Brewer, T.F., Heindel, D.W., Bonneau, R., and Arora, P.S. (2014). Rational Design of Topographical Helix Mimics as Potent Inhibitors of Protein–Protein Interactions. J. Am. Chem. Soc. 136, 7877–7888. DOI: 10.1021/ja502310r

Leach, A.R., and Gillet, V.J. (2007a). Similarity methods. In An Introduction To Chemoinformatics, (Springer Netherlands), pp. 99–117. DOI: 10.1007/978-1-4020-6291-9_5

Leach, A.R., and Gillet, V.J. (2007b). Computational models. In An Introduction To Chemoinformatics, (Springer Netherlands), pp. 75–97. DOI: 10.1007/978-1-4020-6291-9_4

Leach, A.R., and Kuntz, I.D. (1992). Conformational analysis of flexible ligands in macromolecular receptor sites. J. Comput. Chem. 13, 730–748. DOI: 10.1002/jcc.540130608

Leach, A.R., Gillet, V.J., Lewis, R.A., and Taylor, R. (2010). Three-dimensional pharmacophore methods in drug discovery. J. Med. Chem. 53, 539–558. DOI: 10.1021/jm900817u

Lee N.S., Yuen K.Y., and Kumana C.R. (2001). β-lactam antibiotic and β-lactamase inhibitor combinations. JAMA 285, 386–388. DOI: 10.1001/jama.285.4.386

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J.,

Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of Intrinsically Disordered Regions and Proteins. Chem. Rev. 114, 6589–6631. DOI: 10.1021/cr400525m

Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 23, 3–25. DOI: 10.1016/S0169-409X(96)00423-1

Liu, Y., and Gray, N.S. (2006). Rational design of inhibitors that bind to inactive kinase conformations. Nat. Chem. Biol. 2, 358–364. DOI: 10.1038/nchembio799

Lubetsky, J.B., Dios, A., Han, J., Aljabari, B., Ruzsicska, B., Mitchell, R., Lolis, E., and Al-Abed, Y. (2002). The tautomerase active site of macrophage migration inhibitory factor is a potential target for discovery of novel anti-inflammatory agents. J. Biol. Chem. 277, 24976–24982. DOI: 10.1074/jbc.M203220200

Maiti, N.C., Carey, P.R., and Anderson, V.E. (2003). Correlation of an alcohol's αC–D stretch with hydrogen bond strength in complexes with amines. J. Phys. Chem. A 107, 9910–9917. DOI: 10.1021/jp0349334

Maiti, N.C., Zhu, Y., Carmichael, I., Serianni, A.S., and Anderson, V.E. (2006). $^{1}J_{CH}$ correlates with alcohol hydrogen bond strength. J. Org. Chem. 71, 2878–2880. DOI: 10.1021/jo052389k

Maity, M., and Maiti, N.C. (2012). Sequence composition of binding sites in natively unfolded human proteins. J. Proteins Proteomics 3, 117–125.

Maity, M., Pramanik, S.K., Pal, U., Banerji, B., and Maiti, N.C. (2014). Copper(I) oxide nanoparticle and tryptophan as its biological conjugate: a modulation of cytotoxic effects. J. Nanoparticle Res. 16, 1–13. DOI: 10.1007/s11051-013-2179-z

Marshall, T.W. (1961). Isotope shifts in the NMR spectra of H_2, HD and D_2 due to zero-point vibration. Mol. Phys. 4, 61–63. DOI: 10.1080/00268976100100081

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H.J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res. 36, W229–W232. DOI: 10.1093/nar/gkn186

Maxwell, J.C. (1860a). Illustrations of the dynamical theory of gases. Part I. On the motions and collisions of perfectly elastic spheres. Philos. Mag. Ser. 4 19, 19–32. DOI: 10.1080/14786446008642818

Maxwell, J.C. (1860b). Illustrations of the dynamical theory of gases. Part II. On the process of diffusion of two or more kinds of moving particles among one another. Philos. Mag. Ser. 4 20, 21–37. DOI: 10.1080/14786446008642902

McFedries, A., Schwaid, A., and Saghatelian, A. (2013). Methods for the elucidation of protein-small molecule interactions. Chem. Biol. 20, 667–673. DOI: 10.1016/j.chembiol.2013.04.008

McGann, M. (2011). FRED pose prediction and virtual screening accuracy. J. Chem. Inf. Model. 51, 578–596. DOI: 10.1021/ci100436p

Medina-Franco, J.L. (2013). Activity cliffs: facts or artifacts? Chem. Biol. Drug Des. 81, 553–556. DOI: 10.1111/cbdd.12115

Meng, X.-Y., Zhang, H.-X., Mezei, M., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. Curr. Comput. Aided Drug Des. 7, 146–157. DOI: 10.2174/157340911795677602

Mészáros, B., Simon, I., and Dosztányi, Z. (2009). Prediction of protein binding regions in disordered proteins. PLoS Comput Biol 5, e1000376. DOI: 10.1371/journal.pcbi.1000376

Mészáros, B., Dosztányi, Z., and Simon, I. (2012). Disordered binding regions and linear motifs—bridging the gap between two models of molecular recognition. PLoS ONE 7, e46829. DOI: 10.1371/journal.pone.0046829

Metallo, S.J. (2010). Intrinsically disordered proteins are potential drug targets. Curr. Opin. Chem. Biol. 14, 481–488. DOI: 10.1016/j.cbpa.2010.06.169

Millard, C.B., Kryger, G., Ordentlich, A., Greenblatt, H.M., Harel, M., Raves, M.L., Segall, Y., Barak, D., Shafferman, A., Silman, I., et al. (1999). Crystal Structures of Aged Phosphonylated Acetylcholinesterase: Nerve Agent Reaction Products at the Atomic Level. Biochemistry (Mosc.) 38, 7032–7039. DOI: 10.1021/bi982678l

Miller, P.S., and Aricescu, A.R. (2014). Crystal structure of a human $GABA_A$ receptor. Nature 512, 270–275. DOI: 10.1038/nature13293

Mohan, A., Oldfield, C.J., Radivojac, P., Vacic, V., Cortese, M.S., Dunker, A.K., and Uversky, V.N. (2006). Analysis of molecular recognition features (MoRFs). J. Mol. Biol. 362, 1043–1059. DOI: 10.1016/j.jmb.2006.07.087

Monsellier, E., Ramazzotti, M., Taddei, N., and Chiti, F. (2008). Aggregation propensity of the human proteome. PLoS Comput Biol 4, e1000199. DOI: 10.1371/journal.pcbi.1000199

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. 19, 1639–1662. DOI: 10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

Morris, G.M., and Lim-Wilby, M. (2008). Molecular docking. Methods Mol. Biol. Clifton NJ 443, 365–382. DOI: 10.1007/978-1-59745-177-2_19

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–2791. DOI: 10.1002/jcc.21256

Mullard, A. (2012). Drug repurposing programmes get lift off. Nat. Rev. Drug Discov. 11, 505–506. DOI: 10.1038/nrd3776

Munch, M., Hansen, A.E., Hansen, P.E., Bouman, T.D., Abildgaard, F., Led, J.J., and Christensen, S.B. (1992). Ab initio calculations of deuterium isotope effects on hydrogen and nitrogen nuclear magnetic shielding in the hydrated ammonium ion. Acta Chem. Scand. 46, 1065–1071. DOI: 10.3891/acta.chem.scand.46-1065

Murto, J., Kivinen, A., Viitala, R., and Hyömäki, J. (1973). Fluoroalcohols—XX: infrared and Raman spectra of

hexafluoro-2-propanol and its deuterated analogues. Spectrochim. Acta Part Mol. Spectrosc. 29, 1121–1137. DOI: 10.1016/0584-8539(73)80151-5

Nandi, S., Mehra, N., Lynn, A.M., and Bhattacharya, A. (2005). Comparison of theoretical proteomes: Identification of COGs with conserved and variable pI within the multimodal pI distribution. BMC Genomics 6, 116. DOI: 10.1186/1471-2164-6-116

Newman, D.J., and Cragg, G.M. (2007). Natural products as sources of new drugs over the last 25 years. J. Nat. Prod. 70, 461–477. DOI: 10.1021/np068054v

Nguyen Ba, A.N., Yeh, B.J., van Dyk, D., Davidson, A.R., Andrews, B.J., Weiss, E.L., and Moses, A.M. (2012). Proteome-wide discovery of evolutionary conserved sequences in disordered regions. Sci. Signal. 5, rs1. DOI: 10.1126/scisignal.2002515

Nieto, M.M., and Simmons, L.M. (1979). Eigenstates, coherent states, and uncertainty products for the Morse oscillator. Phys. Rev. A 19, 438–444. DOI: 10.1103/PhysRevA.19.438

Nogales, E., Wolf, S.G., and Downing, K.H. (1998). Structure of the αβ tubulin dimer by electron crystallography. Nature 391, 199–203. DOI: 10.1038/34465

Olbe, L., Carlsson, E., and Lindberg, P. (2003). A proton-pump inhibitor expedition: the case histories of omeprazole and esomeprazole. Nat. Rev. Drug Discov. 2, 132–139. DOI: 10.1038/nrd1010

Oldfield, C.J., and Dunker, A.K. (2014). Intrinsically disordered proteins and intrinsically disordered protein regions. Annu. Rev. Biochem. 83, 553–584. DOI: 10.1146/annurev-biochem-072711-164947

Oprea, T.I., Davis, A.M., Teague, S.J., and Leeson, P.D. (2001). Is there a difference between leads and drugs? A historical perspective. J. Chem. Inf. Comput. Sci. 41, 1308–1315. DOI: 10.1021/ci010366a

O'Reilly, M., Vinković, M., Sharff, A., and Jhoti, H. (2006). High throughput protein crystallography: developments in crystallisation, data collection and data processing. Drug Discov. Today Technol. 3, 451–456. DOI: 10.1016/j.ddtec.2006.11.001

Orosz, F., and Ovádi, J. (2011). Proteins without 3D structure: definition, detection and beyond. Bioinforma. Oxf. Engl. 27, 1449–1454. DOI: 10.1093/bioinformatics/btr175

Overington, J.P., Al-Lazikani, B., and Hopkins, A.L. (2006). How many drug targets are there? Nat. Rev. Drug Discov. 5, 993–996. DOI: 10.1038/nrd2199

Pal, U., Maity, M., Khot, N., Das, S., Das, S., Dolui, S., and Maiti, N.C. (2016). Statistical insight into the binding regions in disordered human proteome. J. Proteins Proteomics (In press).

Pal, U., Sen, S., and Maiti, N.C. (2014). C_α–H carries information of a hydrogen bond involving the geminal hydroxyl group: a case study with a hydrogen-bonded complex of 1,1,1,3,3,3-hexafluoro-2-propanol and tertiary amines. J. Phys. Chem. A 118, 1024–1030. DOI: 10.1021/jp411488a

Pal, U., Pramanik, S.K., Bhattacharya, B., Banerji, B., and Maiti, N.C. (2015). Binding interaction of a novel fluorophore with serum albumins: steady state fluorescence perturbation and molecular modeling analysis. SpringerPlus 4, 548. DOI: 10.1186/s40064-015-1333-8

Park, J.-B. (2012). Effects of 17-α ethynyl estradiol on proliferation, differentiation & mineralization of osteoprecursor cells. Indian J. Med. Res. 136, 466–470. PMID: 23041741

Patridge, E., Gareiss, P., Kinch, M.S., and Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. Drug Discov. Today. DOI: 10.1016/j.drudis.2015.01.009

Pellecchia, M., Sem, D.S., and Wüthrich, K. (2002). NMR in drug discovery. Nat. Rev. Drug Discov. 1, 211–219. DOI: 10.1038/nrd748

Pellecchia, M., Bertini, I., Cowburn, D., Dalvit, C., Giralt, E., Jahnke, W., James, T.L., Homans, S.W., Kessler, H., Luchinat, C., et al. (2008). Perspectives on NMR in drug discovery: a technique comes of age. Nat. Rev. Drug Discov. 7, 738–745. DOI: 10.1038/nrd2606

Peng, Z., Yan, J., Fan, X., Mizianty, M.J., Xue, B., Wang, K., Hu, G., Uversky, V.N., and Kurgan, L. (2014). Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. Cell. Mol. Life Sci. 72, 137–151. DOI: 10.1007/s00018-014-1661-9

Perrin, C.L., and Nielson, J.B. (1997). "Strong" hydrogen bonds in chemistry and biology. Annu. Rev. Phys. Chem. 48, 511–544. DOI: 10.1146/annurev.physchem.48.1.511

Petek, B.J., Loggers, E.T., Pollack, S.M., and Jones, R.L. (2015). Trabectedin in soft tissue sarcomas. Mar. Drugs 13, 974–983. DOI: 10.3390/md13020974

Petrone, D., Condemi, J.J., Fife, R., Gluck, O., Cohen, S., and Dalgin, P. (2002). A double-blind, randomized, placebo-controlled study of cevimeline in Sjögren's syndrome patients with xerostomia and keratoconjunctivitis sicca. Arthritis Rheum. 46, 748–754. DOI: 10.1002/art.510

Polinsky, R.J. (1998). Clinical pharmacology of rivastigmine: a new-generation acetylcholinesterase inhibitor for the treatment of alzheimer's disease. Clin. Ther. 20, 634–647. DOI: 10.1016/S0149-2918(98)80127-6

Pryor, R., and Cabreiro, F. (2015). Repurposing metformin: an old drug with new tricks in its binding pockets. Biochem. J. 471, 307–322. DOI: 10.1042/BJ20150497

Rarey, M., Kramer, B., Lengauer, T., and Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 261, 470–489. DOI: 10.1006/jmbi.1996.0477

Ray, A., Seth, B.K., Pal, U., and Basu, S. (2012). Nickel(II)-Schiff base complex recognizing domain II of bovine and human serum albumin: spectroscopic and docking studies. Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 92, 164–174. DOI: 10.1016/j.saa.2012.02.060

Reichenwallner, J., and Hinderberger, D. (2013). Using bound fatty acids to disclose the functional structure of serum albumin. Biochim. Biophys. Acta BBA - Gen. Subj. 1830, 5382–5393. DOI: 10.1016/j.bbagen.2013.04.031

Reymond, J.-L. (2015). The chemical space project. Acc.

Chem. Res. 48, 722–730. DOI: 10.1021/ar500432k

Roberts, R.B., Min, L., Washington, M.K., Olsen, S.J., Settle, S.H., Coffey, R.J., and Threadgill, D.W. (2002). Importance of epidermal growth factor receptor signaling in establishment of adenomas and maintenance of carcinomas during intestinal tumorigenesis. Proc. Natl. Acad. Sci. 99, 1521–1526. DOI: 10.1073/pnas.032678499

Robertson, H.T., and Allison, D.B. (2012). A novel generalized normal distribution for human longevity and other negatively skewed data. PLoS ONE 7. DOI: 10.1371/journal.pone.0037025

Romero, P., Obradovic, Z., Li, X., Garner, E.C., Brown, C.J., and Dunker, A.K. (2001). Sequence complexity of disordered protein. Proteins 42, 38–48. DOI: 10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3

Royer, C.A. (2006). Probing protein folding and conformational transitions with fluorescence. Chem. Rev. 106, 1769–1784. DOI: 10.1021/cr0404390

Rozenberg, M., Shoham, G., Reva, I., and Fausto, R. (2003). Low-temperature Fourier transform infrared spectra and hydrogen bonding in polycrystalline L-alanine. Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 59, 3253–3266. DOI: 10.1016/S1386-1425(03)00159-8

Ruddigkeit, L., van Deursen, R., Blum, L.C., and Reymond, J.-L. (2012). Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. J. Chem. Inf. Model. 52, 2864–2875. DOI: 10.1021/ci300415d

Rudra, D.S., Pal, U., Maiti, N.C., Reiter, R.J., and Swarnakar, S. (2013). Melatonin inhibits matrix metalloproteinase-9 activity by binding to its active site. J. Pineal Res. 54, 398–405. DOI: 10.1111/jpi.12034

Rupakheti, C., Virshup, A., Yang, W., and Beratan, D.N. (2015). Strategy to discover diverse optimal molecules in the small molecule universe. J. Chem. Inf. Model. 55, 529–537. DOI: 10.1021/ci500749q

Sandler, A., Gray, R., Perry, M.C., Brahmer, J., Schiller, J.H., Dowlati, A., Lilenbaum, R., and Johnson, D.H. (2006). Paclitaxel–carboplatin alone or with bevacizumab for non–small-cell lung cancer. N. Engl. J. Med. 355, 2542–2550. DOI: 10.1056/NEJMoa061884

Sanson, B., Nachon, F., Colletier, J.-P., Froment, M.-T., Toker, L., Greenblatt, H.M., Sussman, J.L., Ashani, Y., Masson, P., Silman, I., et al. (2009). Crystallographic Snapshots of Nonaged and Aged Conjugates of Soman with Acetylcholinesterase, and of a Ternary Complex of the Aged Conjugate with Pralidoxime†. J. Med. Chem. 52, 7593–7603. DOI: 10.1021/jm900433t

Schad, E., Kalmar, L., and Tompa, P. (2013). Exon-phase symmetry and intrinsic structural disorder promote modular evolution in the human genome. Nucleic Acids Res. 41, 4409–4422. DOI: 10.1093/nar/gkt110

Schamberger, J., Grimm, M., Steinmeyer, A., and Hillisch, A. (2011). Rendezvous in chemical space? Comparing the small molecule compound libraries of Bayer and Schering. Drug Discov. Today 16, 636–641. DOI: 10.1016/j.drudis.2011.04.005

Schauer, G.D., Huber, K.D., Leuba, S.H., and Sluis-Cremer, N. (2014). Mechanism of allosteric inhibition of HIV-1 reverse transcriptase revealed by single-molecule and ensemble fluorescence. Nucleic Acids Res. 42, 11687–11696. DOI: 10.1093/nar/gku819

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. 33, W363–W367. DOI: 10.1093/nar/gki481

Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., et al. (2013). BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. 41, D764–D772. DOI: 10.1093/nar/gks1049

Schramm, V.L. (2005). Enzymatic transition states and transition state analogues. Curr. Opin. Struct. Biol. 15, 604–613. DOI: 10.1016/j.sbi.2005.10.017

Schramm, V.L. (2007). Enzymatic transition state theory and transition state analogue design. J. Biol. Chem. 282, 28297–28300. DOI: 10.1074/jbc.R700018200

Schramm, V.L. (2013). Transition states, analogues and drug development. ACS Chem. Biol. 8, 71–81. DOI: 10.1021/cb300631k

Schwartz, T.W., and Holst, B. (2007). Allosteric enhancers, allosteric agonists and ago-allosteric modulators: where do they bind and how do they act? Trends Pharmacol. Sci. 28, 366–373. DOI: 10.1016/j.tips.2007.06.008

Seckler, J.M., Barkley, M.D., and Wintrode, P.L. (2011). Allosteric suppression of HIV-1 reverse transcriptase structural dynamics upon inhibitor binding. Biophys. J. 100, 144–153. DOI: 10.1016/j.bpj.2010.11.004

Shanbhag, D.N., and Rao, C.R. (2003). Stochastic Processes: Modelling and Simulation (Gulf Professional Publishing). ISBN 13: 9780444500137

Sharma, A., Dehzangi, A., Lyons, J., Imoto, S., Miyano, S., Nakai, K., and Patil, A. (2014). Evaluation of Sequence Features from Intrinsically Disordered Regions for the Estimation of Protein Function. PLoS ONE 9, e89890. DOI: 10.1371/journal.pone.0089890

Sharp, D.J., Rogers, G.C., and Scholey, J.M. (2000). Microtubule motors in mitosis. Nature 407, 41–47. DOI: 10.1038/35024000

Sickmeier, M., Hamilton, J.A., LeGall, T., Vacic, V., Cortese, M.S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V.N., et al. (2007). DisProt: the database of disordered proteins. Nucleic Acids Res. 35, D786–D793. DOI: 10.1093/nar/gkl893

Silverman, R.B., and Holladay, M.W. (2014). Chapter 2 — Lead discovery and lead modification. In The Organic Chemistry of Drug Design and Drug Action (Third Edition), (Boston: Academic Press), pp. 19–122. DOI: 10.1016/B978-0-12-382030-3.00002-7

Simard, J.R., Zunszain, P.A., Hamilton, J.A., and Curry, S. (2006). Location of high and low affinity fatty acid binding sites on human serum albumin revealed by NMR drug-competition analysis. J. Mol. Biol. 361, 336–351. DOI: 10.1016/j.jmb.2006.06.028

Skinner, A.L., and Laurence, J.S. (2008). High-field solution NMR spectroscopy as a tool for assessing protein interactions

with small molecule ligands. J. Pharm. Sci. 97, 4670–4695. DOI: 10.1002/jps.21378

Sneader, W. (2000). The discovery of aspirin: a reappraisal. BMJ 321, 1591–1594. DOI: 10.1136/bmj.321.7276.1591

Sousa, S.F., Fernandes, P.A., and Ramos, M.J. (2006). Protein-ligand docking: current status and future challenges. Proteins 65, 15–26. DOI: 10.1002/prot.21082

Stanton, R.A., Gernert, K.M., Nettles, J.H., and Aneja, R. (2011). Drugs that target dynamic microtubules: a new molecular perspective. Med. Res. Rev. 31, 443–481. DOI: 10.1002/med.20242

Sturlese, M., Bellanda, M., and Moro, S. (2015). NMR-assisted molecular docking methodologies. Mol. Inform. 34, 513–525. DOI: 10.1002/minf.201500012

Sugase, K., Dyson, H.J., and Wright, P.E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447, 1021–1025. DOI: 10.1038/nature05858

Tao, P., Fisher, J.F., Shi, Q., Vreven, T., Mobashery, S., and Schlegel, H.B. (2009). Matrix metalloproteinase 2 inhibition: combined quantum mechanics and molecular mechanics studies of the inhibition mechanism of (4-phenoxyphenylsulfonyl)methylthiirane and its oxirane analogue. Biochemistry 48, 9839–9847. DOI: 10.1021/bi901118r

Taylor, J.P., Hardy, J., and Fischbeck, K.H. (2002). Toxic proteins in neurodegenerative disease. Science 296, 1991–1995. DOI: 10.1126/science.1067122

Taylor, R.D., Jewsbury, P.J., and Essex, J.W. (2002). A review of protein-small molecule docking methods. J. Comput. Aided Mol. Des. 16, 151–166. DOI: 10.1023/A:1020155510718

Tetko, I.V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V.A., Radchenko, E.V., Zefirov, N.S., Makarenko, A.S., et al. (2005). Virtual computational chemistry laboratory--design and description. J. Comput. Aided Mol. Des. 19, 453–463. DOI: 10.1007/s10822-005-8694-y

Tompa, P., and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperones. FASEB J. 18, 1169–1175. DOI: 10.1096/fj.04-1584rev

Tonge, P.J., Fausto, R., and Carey, P.R. (1996). FTIR studies of hydrogen bonding between α,β-unsaturated esters and alcohols. J. Mol. Struct. 379, 135–142.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. DOI: 10.1002/jcc.21334

Tsvetkov, P., Reuven, N., and Shaul, Y. (2009). The nanny model for IDPs. Nat. Chem. Biol. 5, 778–781. DOI: 10.1038/nchembio.233

Uversky, V.N. (2002). What does it mean to be natively unfolded? Eur. J. Biochem. FEBS 269, 2–12. DOI: 10.1046/j.0014-2956.2001.02649.x

Uversky, V.N., Gillespie, J.R., and Fink, A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41, 415–427. DOI: 10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7

Vacic, V., Oldfield, C.J., Mohan, A., Radivojac, P., Cortese, M.S., Uversky, V.N., and Dunker, A.K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. J. Proteome Res. 6, 2351–2366. DOI: 10.1021/pr0701411

Venkatachalam, C.M., Jiang, X., Oldfield, T., and Waldman, M. (2003). LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J. Mol. Graph. Model. 21, 289–307. DOI: 10.1016/S1093-3263(02)00164-X

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., and Taylor, R.D. (2003). Improved protein–ligand docking using GOLD. Proteins Struct. Funct. Bioinforma. 52, 609–623. DOI: 10.1002/prot.10465

Virshup, A.M., Contreras-García, J., Wipf, P., Yang, W., and Beratan, D.N. (2013). Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. J. Am. Chem. Soc. 135, 7296–7303. DOI: 10.1021/ja401184g

Volkamer, A., Kuhn, D., Rippmann, F., and Rarey, M. (2012). DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. Bioinformatics 28, 2074–2075. DOI: 10.1093/bioinformatics/bts310

Vucetic, S., Brown, C.J., Dunker, A.K., and Obradovic, Z. (2003). Flavors of protein disorder. Proteins 52, 573–584. DOI: 10.1002/prot.10437

Wang, D., and Lippard, S.J. (2005). Cellular processing of platinum anticancer drugs. Nat. Rev. Drug Discov. 4, 307–320. DOI: 10.1038/nrd1691

Warfield, R.K., and Bouck, G.B. (1974). Microtubule-macrotubule transitions: intermediates after exposure to the mitotic inhibitor vinblastine. Science 186, 1219–1221. DOI: 10.1126/science.186.4170.1219

Weinreb, P.H., Zhen, W., Poon, A.W., Conway, K.A., and Lansbury, P.T., Jr (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. Biochemistry 35, 13709–13715. DOI: 10.1021/bi961799n

Wendler, K., Thar, J., Zahn, S., and Kirchner, B. (2010). Estimating the hydrogen bond energy. J. Phys. Chem. A 114, 9529–9536. DOI: 10.1021/jp103470e

Wermuth, C.G., Ganellin, C.R., Lindberg, P., and Mitscher, L.A. (2009). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). Pure Appl. Chem. 70, 1129–1143. DOI: 10.1351/pac199870051129

White, N.J. (2008). Qinghaosu (Artemisinin): The Price of Success. Science 320, 330–334. DOI: 10.1126/science.1155165

Willett, P. (2006). Similarity-based virtual screening using 2D fingerprints. Drug Discov. Today 11, 1046–1053. DOI: 10.1016/j.drudis.2006.10.005

Woestenenk, E.A., Hammarström, M., Härd, T., and Berglund, H. (2003). Screening methods to determine biophysical properties of proteins in structural genomics. Anal. Biochem. 318, 71–79. DOI: 10.1016/S0003-2697(03)00162-3

Wolinski, K., Hinton, J.F., and Pulay, P. (1990). Efficient implementation of the gauge-independent atomic orbital

method for NMR chemical shift calculations. J. Am. Chem. Soc. 112, 8251–8260. DOI: 10.1021/ja00179a005

Woody, R.W. (1996). Theory of circular dichroism of proteins. In Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, ed. (Springer US), pp. 25–67. DOI: 10.1007/978-1-4757-2508-7_2

Wright, P.E., and Dyson, H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. 293, 321–331. DOI: 10.1006/jmbi.1999.3110

Wu, S., Wan, P., Li, J., Li, D., Zhu, Y., and He, F. (2006). Multi-modality of pI distribution in whole proteome. PROTEOMICS 6, 449–455. DOI: 10.1002/pmic.200500221

Xiao, H., Verdier-Pinard, P., Fernandez-Fuentes, N., Burd, B., Angeletti, R., Fiser, A., Horwitz, S.B., and Orr, G.A. (2006). Insights into the mechanism of microtubule stabilization by Taxol. Proc. Natl. Acad. Sci. 103, 10166–10173. DOI: 10.1073/pnas.0603704103

Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N., and Obradovic, Z. (2007a). Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J. Proteome Res. 6, 1882–1898. DOI: 10.1021/pr060392u

Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Obradovic, Z., and Uversky, V.N. (2007b). Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. J. Proteome Res. 6, 1917–1932. DOI: 10.1021/pr060394e

Yamasaki, K., Chuang, V.T.G., Maruyama, T., and Otagiri, M. (2013). Albumin–drug interaction and its clinical implication. Biochim. Biophys. Acta BBA - Gen. Subj. 1830, 5435–5443. DOI: 10.1016/j.bbagen.2013.05.005

Zhang, J., Adrián, F.J., Jahnke, W., Cowan-Jacob, S.W., Li, A.G., Iacob, R.E., Sim, T., Powers, J., Dierks, C., Sun, F., et al. (2010). Targeting Bcr–Abl by combining allosteric with ATP-binding-site inhibitors. Nature 463, 501–506. DOI: 10.1038/nature08675

Zhou, Z., Zhen, J., Karpowich, N.K., Goetz, R.M., Law, C.J., Reith, M.E.A., and Wang, D.-N. (2007). LeuT-desipramine structure reveals how antidepressants block neurotransmitter reuptake. Science 317, 1390–1393. DOI: 10.1126/science.1147614

APPENDIX I

Refers to: Chapter 3

"If this is victory, then our hands are too small to hold it."

J.R.R. Tolkien, The Return of the King

List of the disordered proteins in the dataset

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
1	P49913	Cathelicidin antimicrobial peptide	0.00	0
2	P27695	DNA-(apurinic or apyrimidinic site) lyase	23.03	1
3	P50224	Sulfotransferase 1A3/1A4	5.42	1
4	P13569	Cystic fibrosis transmembrane conductance regulator	3.92	2
5	P01233	Choriogonadotropin subunit beta	24.14	1
6	P38936	Cyclin-dependent kinase inhibitor 1	47.85	5
7	P49918	Cyclin-dependent kinase inhibitor 1C	65.19	5
8	P46527	Cyclin-dependent kinase inhibitor 1B	95.45	8
9	P14061	Estradiol 17-beta-dehydrogenase 1	13.15	3
10	Q13541	Eukaryotic translation initiation factor 4E-binding protein 1	96.58	5
11	P04150	Glucocorticoid receptor	31.92	10
12	P10912	Growth hormone receptor	27.10	8
13	P05204	Non-histone chromosomal protein HMG-17	100.00	3
14	P17096	High mobility group protein HMG-I/HMG-Y	100.00	3
15	P78356	Phosphatidylinositol 5-phosphate 4-kinase type-2 beta	12.53	4
16	P01106	Myc proto-oncogene protein	47.84	9
17	P27694	Replication protein A 70 kDa DNA-binding subunit	8.28	5
18	P19793	Retinoic acid receptor RXR-alpha	32.68	3
19	P63027	Vesicle-associated membrane protein 2	28.70	1
20	P04818	Thymidylate synthase	8.97	1
21	P03372	Estrogen receptor	19.50	1
22	P11387	DNA topoisomerase 1	31.54	8
23	P01100	Proto-oncogene c-Fos	50.00	8
24	P61244	Protein max	100.00	5
25	Q04206	Transcription factor p65	56.81	12
26	P04637	Cellular tumor antigen p53	53.18	8
27	Q08209	Serine/threonine-protein phosphatase 2B catalytic subunit	15.58	2
28	P63098	Calcineurin subunit B type 1	0.00	0
29	P50440	Glycine amidinotransferase, mitochondrial	5.18	0
30	P04350	Tubulin beta-4A chain	7.43	0
31	P10163	Basic salivary proline-rich protein 4	100.00	12
32	P01730	T-cell surface glycoprotein CD4	5.54	1
33	P22531	Small proline-rich protein 2E	0.00	0

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
34	Q06787	Fragile X mental retardation protein 1	41.14	10
35	P12272	Parathyroid hormone-related protein	89.36	5
36	Q13426	DNA repair protein XRCC4	40.18	5
37	P01112	GTPase HRas	3.23	1
38	P29353	SHC-transforming protein 1	57.12	10
39	P54725	UV excision repair protein RAD23 homolog A	60.06	8
40	O75324	Stannin	11.36	0
41	P19429	Troponin I, cardiac muscle	35.89	4
42	P10114	Ras-related protein Rap-2a	1.11	0
43	Q9NVS9	Pyridoxine-5'-phosphate oxidase	9.58	2
44	P04049	RAF proto-oncogene serine/threonine-protein kinase	20.68	3
45	Q16633	POU domain class 2-associating factor 1	41.80	3
46	P16949	Stathmin	91.89	4
47	Q9NQB0	Transcription factor 7-like 2	68.17	13
48	Q9Y5B0	RNA polymerase II subunit A C-terminal domain phosphatase	57.86	17
49	P11473	Vitamin D3 receptor	14.52	2
50	P00747	Plasminogen	13.78	2
51	P20810	Calpastatin	100.00	20
52	P05814	Beta-casein	20.38	1
53	P62993	Growth factor receptor-bound protein 2	0.00	0
54	P24592	Insulin-like growth factor-binding protein 6	60.09	2
55	P10451	Osteopontin	98.66	6
56	P42768	Wiskott-Aldrich syndrome protein	76.05	6
57	O60927	Protein phosphatase 1 regulatory subunit 11	84.80	3
58	P07476	Involucrin	99.49	11
59	P14635	G2/mitotic-specific cyclin-B1	30.72	5
60	P15884	Transcription factor 4	95.95	18
61	P05455	Lupus La protein	34.07	2
62	Q14449	Growth factor receptor-bound protein 14	11.32	3
63	P14859	POU domain, class 2, transcription factor 1	39.97	9
64	P02686	Myelin basic protein	93.42	10
65	P38398	Breast cancer type 1 susceptibility protein	62.05	44
66	P23025	DNA repair protein complementing XP-A cells	16.18	3
67	Q93096	Protein tyrosine phosphatase type IVA 1	0.00	0
68	Q9Y3R4	Sialidase-2	8.68	0
69	P40337	Von Hippel-Lindau disease tumor suppressor	41.78	2
70	P10415	Apoptosis regulator Bcl-2	23.85	2
71	Q07817	Bcl-2-like protein 1	17.60	1
72	P31751	RAC-beta serine/threonine-protein kinase	8.73	1
73	Q07108	Early activation antigen CD69	6.53	0
74	P51946	Cyclin-H	5.26	0
75	P06132	Uroporphyrinogen decarboxylase	4.90	0
76	P62495	Eukaryotic peptide chain release factor subunit 1	2.52	0
77	P36888	Receptor-type tyrosine-protein kinase FLT3	2.17	1
78	P32455	Interferon-induced guanylate-binding protein 1	15.11	1
79	Q8WUM0	Nuclear pore complex protein Nup133	5.97	3
80	P05121	Plasminogen activator inhibitor 1	0.00	0
81	Q9UM07	Protein-arginine deiminase type-4	5.88	2
82	O14832	Phytanoyl-CoA dioxygenase, peroxisomal	4.55	0
83	P21815	Bone sialoprotein 2	91.03	7

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
84	P27797	Calreticulin	54.00	5
85	P16442	Histo-blood group ABO system transferase	1.41	0
86	P02649	Apolipoprotein E	18.06	1
87	P62328	Thymosin beta-4	90.70	1
88	Q15185	Prostaglandin E synthase 3	26.88	3
89	Q15382	GTP-binding protein Rheb	1.66	0
90	P53041	Serine/threonine-protein phosphatase 5	8.43	0
91	Q9NR00	Uncharacterized protein C8orf4	3.77	0
92	P08047	Transcription factor Sp1	54.97	14
93	P35869	Aryl hydrocarbon receptor	24.82	8
94	P05091	Aldehyde dehydrogenase, mitochondrial	4.80	1
95	P02647	Apolipoprotein A-I	28.92	2
96	Q09161	Nuclear cap-binding protein subunit 1	6.84	1
97	O00204	Sulfotransferase family cytosolic 2B member 1	21.37	2
98	Q04637	Eukaryotic translation initiation factor 4 gamma 1	60.29	27
99	P78504	Protein jagged-1	10.97	6
100	Q96T21	Selenocysteine insertion sequence-binding protein 2	53.86	15
101	P53350	Serine/threonine-protein kinase PLK1	11.96	3
102	P43351	DNA repair protein RAD52 homolog	55.98	7
103	Q14191	Werner syndrome ATP-dependent helicase	15.86	10
104	P02489	Alpha-crystallin A chain	12.14	1
105	P02511	Alpha-crystallin B chain	22.86	1
106	Q9Y613	FH1/FH2 domain-containing protein 1	43.34	15
107	Q9Y3D6	Mitochondrial fission 1 protein	1.97	0
108	Q07960	Rho GTPase-activating protein 1	13.21	2
109	Q92838	Ectodysplasin-A	54.22	5
110	Q13485	Mothers against decapentaplegic homolog 4	28.26	4
111	P04156	Major prion protein	52.88	5
112	Q01484	Ankyrin-2	59.64	76
113	P25963	NF-kappa-B inhibitor alpha	26.81	3
114	Q49AN0	Cryptochrome-2	11.97	1
115	Q9UBE0	SUMO-activating enzyme subunit 1	10.69	0
116	Q9UBT2	SUMO-activating enzyme subunit 2	29.22	5
117	P52434	DNA-directed RNA polymerases I, II, and III subunit RPABC3	6.71	0
118	P04234	T-cell surface glycoprotein CD3 delta chain	0.00	0
119	P07766	T-cell surface glycoprotein CD3 epsilon chain	30.81	2
120	P09693	T-cell surface glycoprotein CD3 gamma chain	14.38	0
121	P30273	High affinity immunoglobulin epsilon receptor subunit gamma	17.65	0
122	O60356	Nuclear protein 1	100.00	3
123	O95997	Securin	52.24	6
124	P51608	Methyl-CpG-binding protein 2	95.47	12
125	Q14061	Cytochrome c oxidase copper chaperone	19.05	0
126	Q9NX55	Huntingtin-interacting protein K	86.82	4
127	O00488	Zinc finger protein 593	85.82	4
128	P16860	Natriuretic peptides B	79.63	4
129	Q16143	Beta-synuclein	70.15	3
130	Q9Y296	Trafficking protein particle complex subunit 4	0.00	0
131	P09132	Signal recognition particle 19 kDa protein	23.61	2
132	P39900	Macrophage metalloelastase	4.38	0
133	P11831	Serum response factor	71.85	12

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
134	P11233	Ras-related protein Ral-A	10.34	2
135	P48539	Purkinje cell protein 4	100.00	2
136	Q16595	Frataxin, mitochondrial	7.10	0
137	Q13573	SNW domain-containing protein 1	67.85	17
138	P30291	Wee1-like protein kinase	46.13	8
139	P02788	Lactotransferrin	0.28	0
140	P60896	26S proteasome complex subunit DSS1	97.14	1
141	P00736	Complement C1r subcomponent	1.16	0
142	P35638	DNA damage-inducible transcript 3 protein	88.76	6
143	O76070	Gamma-synuclein	70.87	3
144	Q01844	RNA-binding protein EWS	90.55	14
145	Q09472	Histone acetyltransferase p300	67.68	29
146	P01270	Parathyroid hormone	55.95	3
147	P51671	Eotaxin	18.92	1
148	O43521	Bcl-2-like protein 11	49.49	2
149	P22362	C-C motif chemokine 1	0.00	0
150	P31025	Lipocalin-1	18.99	0
151	P00441	Superoxide dismutase [Cu-Zn]	35.29	5
152	Q29983	MHC class I polypeptide-related sequence A	11.67	1
153	Q92886	Neurogenin-1	58.65	4
154	P11171	Protein 4.1	60.42	19
155	Q99801	Homeobox protein Nkx-3.1	59.83	3
156	P18754	Regulator of chromosome condensation	18.57	4
157	Q8N488	RING1 and YY1-binding protein	89.91	5
158	Q9Y258	C-C motif chemokine 26	0.00	0
159	P46937	Yorkie homolog	92.86	14
160	P24522	Growth arrest and DNA damage-inducible protein GADD45 alpha	8.48	0
161	Q15517	Corneodesmosin	74.25	11
162	Q9UER7	Death domain-associated protein 6	64.59	18
163	Q9Y3M2	Protein chibby homolog 1	40.48	2
164	Q9UGL1	Lysine-specific demethylase 5B	16.65	5
165	P29375	Lysine-specific demethylase 5A	18.58	10
166	P02545	Prelamin-A/C	52.56	13
167	P37231	Peroxisome proliferator-activated receptor gamma	6.14	0
168	Q13569	G/T mismatch-specific thymine DNA glycosylase	28.54	5
169	P16083	Ribosyldihydronicotinamide dehydrogenase [quinone]	0.00	0
170	P07602	Prosaposin	2.17	1
171	Q9UKV8	Protein argonaute-2	9.66	2
172	P01160	Natriuretic peptides A	16.67	1
173	P01160	Natriuretic peptides A	0.00	0
174	P46108	Adapter molecule crk	41.25	8
175	Q9BRL6	Serine/arginine-rich splicing factor 8	90.39	7
176	P37840	Alpha-synuclein	40.00	2
177	O00482	Nuclear receptor subfamily 5 group A member 2	14.42	3
178	O00512	B-cell CLL/lymphoma 9 protein	99.86	22
179	O14646	Chromodomain-helicase-DNA-binding protein 1	49.59	32
180	O14717	tRNA (cytosine(38)-C(5))-methyltransferase	0.77	0
181	O15151	Protein Mdm4	31.22	4
182	O15162	Phospholipid scramblase 1	22.96	4
183	O15169	Axin-1	64.15	17

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
184	O60260	E3 ubiquitin-protein ligase parkin	9.68	1
185	O60341	Lysine-specific histone demethylase 1A	27.11	8
186	O60814	Histone H2B type 1-K	33.60	2
187	O95718	Steroid hormone receptor ERR2	16.93	5
188	P04908	Histone H2A type 1-B/E	25.58	0
189	P06400	Retinoblastoma-associated protein	22.22	7
190	P09012	U1 small nuclear ribonucleoprotein A	45.91	6
191	P0C0S5	Histone H2A.Z	30.71	1
192	P10275	Androgen receptor	38.41	12
193	P11474	Steroid hormone receptor ERR1	15.60	2
194	P12004	Proliferating cell nuclear antigen	3.83	0
195	P12956	X-ray repair cross-complementing protein 6	12.83	4
196	P13010	X-ray repair cross-complementing protein 5	6.98	2
197	P14921	Protein C-ets-1	4.76	0
198	P15927	Replication protein A 32 kDa subunit	28.89	2
199	P16104	Histone H2AX	30.28	0
200	P16401	Histone H1.5	84.00	6
201	P17931	Galectin-3	42.17	3
202	P18615	Negative elongation factor E	73.95	8
203	P20248	Cyclin-A2	33.80	4
204	P20393	Nuclear receptor subfamily 1 group D member 1	51.63	9
205	P24941	Cyclin-dependent kinase 2	0.00	0
206	P25054	Adenomatous polyposis coli protein	75.83	55
207	P31946	14-3-3 protein beta/alpha	24.80	2
208	P35222	Catenin beta-1	24.10	8
209	P35244	Replication protein A 14 kDa subunit	0.83	0
210	P39748	Flap endonuclease 1	19.74	4
211	P41235	Hepatocyte nuclear factor 4-alpha	20.04	3
212	P42224	Signal transducer and activator of transcription 1-alpha/beta	7.61	4
213	P42226	Signal transducer and activator of transcription 6	28.49	8
214	P43246	DNA mismatch repair protein Msh2	0.54	1
215	P48552	Nuclear receptor-interacting protein 1	64.42	31
216	P49005	DNA polymerase delta subunit 2	4.48	1
217	P49841	Glycogen synthase kinase-3 beta	27.14	7
218	P52630	Signal transducer and activator of transcription 2	10.81	5
219	P52701	DNA mismatch repair protein Msh6	20.74	9
220	P62158	Calmodulin	43.92	4
221	P62508	Estrogen-related receptor gamma	8.52	1
222	P62805	Histone H4	18.63	0
223	P62877	E3 ubiquitin-protein ligase RBX1	1.85	0
224	P63104	14-3-3 protein zeta/delta	12.65	1
225	P68431	Histone H3.1	33.33	1
226	Q00688	Peptidyl-prolyl cis-trans isomerase FKBP3	26.01	4
227	Q01094	Transcription factor E2F1	45.08	10
228	Q01167	Forkhead box protein K2	64.95	18
229	Q02156	Protein kinase C epsilon type	15.60	4
230	Q02548	Paired box protein Pax-5	56.78	9
231	Q02790	Peptidyl-prolyl cis-trans isomerase FKBP4	16.34	2
232	Q12830	Nucleosome-remodeling factor subunit BPTF	61.69	56
233	Q13285	Steroidogenic factor 1	22.56	5

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
234	Q13451	Peptidyl-prolyl cis-trans isomerase FKBP5	18.82	3
235	Q13772	Nuclear receptor coactivator 4	21.34	10
236	Q14103	Heterogeneous nuclear ribonucleoprotein D0	43.38	6
237	Q14186	Transcription factor Dp-1	36.19	7
238	Q14541	Hepatocyte nuclear factor 4-gamma	13.73	2
239	Q14565	Meiotic recombination protein DMC1/LIM15 homolog	6.76	1
240	Q14653	Interferon regulatory factor 3	35.13	4
241	Q15054	DNA polymerase delta subunit 3	67.74	11
242	Q15466	Nuclear receptor subfamily 0 group B member 2	3.11	0
243	Q15596	Nuclear receptor coactivator 2	82.30	29
244	Q15788	Nuclear receptor coactivator 1	72.08	37
245	Q16576	Histone-binding protein RBBP7	16.98	3
246	Q16665	Hypoxia-inducible factor 1-alpha	42.25	12
247	Q16695	Histone H3.1t	33.33	1
248	Q6W2J9	BCL-6 corepressor	61.03	38
249	Q71DI3	Histone H3.2	33.33	1
250	Q86VP6	Cullin-associated NEDD8-dissociated protein 1	3.82	3
251	Q92540	Protein SMG7	50.79	16
252	Q92731	Estrogen receptor beta	10.38	2
253	Q92793	CREB-binding protein	68.00	34
254	Q92837	Proto-oncogene FRAT1	85.30	8
255	Q92973	Transportin-1	4.34	2
256	Q969R5	Lethal(3)malignant brain tumor-like protein 2	24.54	5
257	Q96AY2	Crossover junction endonuclease EME1	40.88	8
258	Q96EP1	E3 ubiquitin-protein ligase CHFR	43.22	9
259	Q96NY9	Crossover junction endonuclease MUS81	28.49	9
260	Q99741	Cell division control protein 6 homolog	24.29	5
261	Q9NQR1	N-lysine methyltransferase SETD8	50.38	7
262	Q9NSA3	Beta-catenin-interacting protein 1	51.85	3
263	Q9UBK2	Peroxisome proliferator-activated receptor gamma coactivator	61.90	17
264	Q9UBT6	DNA polymerase kappa	13.79	8
265	Q9UNA4	DNA polymerase iota	25.54	8
266	Q9UPR3	Protein SMG5	23.05	3
267	Q9Y253	DNA polymerase eta	27.21	10
268	Q9Y618	Nuclear receptor corepressor 2	86.22	55
269	O60885	Bromodomain-containing protein 4	84.73	31
270	O95405	Zinc finger FYVE domain-containing protein 9	30.95	20
271	P84022	Mothers against decapentaplegic homolog 3	17.92	1
272	Q96PU5	E3 ubiquitin-protein ligase NEDD4-like	42.40	13
273	O95149	Snurportin-1	29.72	4
274	Q9H816	5' exonuclease Apollo	16.73	3
275	O76064	E3 ubiquitin-protein ligase RNF8	31.13	5
276	P09651	Heterogeneous nuclear ribonucleoprotein A1	54.03	8
277	Q8IUQ4	E3 ubiquitin-protein ligase SIAH1	1.77	0
278	P22415	Upstream stimulatory factor 1	65.16	7
279	P52907	F-actin-capping protein subunit alpha-1	5.96	0
280	Q12888	Tumor suppressor p53-binding protein 1	79.61	41
281	Q13625	Apoptosis-stimulating of p53 protein 2	66.76	19
282	Q9HB71	Calcyclin-binding protein	18.06	2
283	P24928	DNA-directed RNA polymerase II subunit RPB1	31.22	5

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
284	Q15796	Mothers against decapentaplegic homolog 2	18.88	1
285	Q8WTS6	Histone-lysine N-methyltransferase SETD7	10.11	1
286	O43707	Alpha-actinin-4	8.34	3
287	P28749	Retinoblastoma-like protein 1	17.04	8
288	P12755	Ski oncogene	40.93	10
289	Q13616	Cullin-1	5.15	1
290	O14745	Na(+)/H(+) exchange regulatory cofactor NHE-RF1	75.35	11
291	O15105	Mothers against decapentaplegic homolog 7	21.36	3
292	O75928	E3 SUMO-protein ligase PIAS2	19.32	4
293	P04083	Annexin A1	7.83	0
294	P08621	U1 small nuclear ribonucleoprotein 70 kDa	83.26	8
295	P09234	U1 small nuclear ribonucleoprotein C	87.42	2
296	P14653	Homeobox protein Hox-B1	76.08	8
297	P14678	Small nuclear ribonucleoprotein-associated proteins B and B'	71.67	6
298	P19419	ETS domain-containing protein Elk-1	68.93	8
299	P26368	Splicing factor U2AF 65 kDa subunit	31.22	3
300	P28324	ETS domain-containing protein Elk-4	48.26	5
301	P40424	Pre-B-cell leukemia transcription factor 1	50.93	8
302	P47974	Zinc finger protein 36, C3H1 type-like 2	48.58	9
303	P49792	E3 SUMO-protein ligase RanBP2	27.79	33
304	P52272	Heterogeneous nuclear ribonucleoprotein M	34.43	7
305	P52292	Importin subunit alpha-1	18.75	5
306	P54198	Protein HIRA	20.55	7
307	P54274	Telomeric repeat-binding factor 1	36.07	4
308	P62304	Small nuclear ribonucleoprotein E	0.00	0
309	P62306	Small nuclear ribonucleoprotein F	7.06	0
310	P62308	Small nuclear ribonucleoprotein G	2.63	0
311	P62314	Small nuclear ribonucleoprotein Sm D1	37.82	1
312	P62316	Small nuclear ribonucleoprotein Sm D2	37.61	2
313	P62318	Small nuclear ribonucleoprotein Sm D3	19.20	3
314	P62826	GTP-binding nuclear protein Ran	3.26	0
315	Q00987	E3 ubiquitin-protein ligase Mdm2	47.45	8
316	Q01831	DNA repair protein complementing XP-C cells	45.26	15
317	Q05195	Max dimerization protein 1	50.68	7
318	Q05639	Elongation factor 1-alpha 2	13.17	3
319	Q13043	Serine/threonine-protein kinase 4	35.73	5
320	Q13127	RE1-silencing transcription factor	67.55	22
321	Q14209	Transcription factor E2F2	49.20	10
322	Q15208	Serine/threonine-protein kinase 38	10.34	2
323	Q15648	Mediator of RNA polymerase II transcription subunit 1	66.60	28
324	Q15797	Mothers against decapentaplegic homolog 1	27.53	4
325	Q93009	Ubiquitin carboxyl-terminal hydrolase 7	10.80	4
326	Q96IZ0	PRKC apoptosis WT1 regulator protein	83.53	7
327	Q96J02	E3 ubiquitin-protein ligase Itchy homolog	25.72	12
328	Q99967	Cbp/p300-interacting transactivator 2	77.78	3
329	Q9BSI4	TERF1-interacting nuclear factor 2	34.67	8
330	Q9BX63	Fanconi anemia group J protein	18.65	10
331	Q9Y294	Histone chaperone ASF1A	17.16	2
332	Q9Y6J0	Calcineurin-binding protein cabin-1	37.39	31
333	O14777	Kinetochore protein NDC80 homolog	21.96	4

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
334	O15234	Protein CASC3	97.72	15
335	O15392	Baculoviral IAP repeat-containing protein 5	14.08	2
336	O60828	Polyglutamine-binding protein 1	81.82	5
337	O75362	Zinc finger protein 217	57.63	20
338	O75376	Nuclear receptor corepressor 1	87.83	62
339	O75533	Splicing factor 3B subunit 1	32.29	13
340	O95071	E3 ubiquitin-protein ligase UBR5	34.92	30
341	O95633	Follistatin-related protein 3	8.86	0
342	P11940	Polyadenylate-binding protein 1	32.08	8
343	P15941	Mucin-1	83.04	4
344	P26367	Paired box protein Pax-6	61.37	10
345	P30307	M-phase inducer phosphatase 3	33.90	6
346	P46531	Neurogenic locus notch homolog protein 1	21.60	18
347	P55072	Transitional endoplasmic reticulum ATPase	12.67	6
348	P56915	Homeobox protein goosecoid	23.74	0
349	P84103	Serine/arginine-rich splicing factor 3	56.10	4
350	Q01081	Splicing factor U2AF 35 kDa subunit	24.27	1
351	Q01658	Protein Dr1	17.71	1
352	Q04724	Transducin-like enhancer protein 1	40.78	9
353	Q05086	Ubiquitin-protein ligase E3A	10.97	2
354	Q12802	A-kinase anchor protein 13	61.75	56
355	Q14469	Transcription factor HES-1	62.86	7
356	Q15599	Na(+)/H(+) exchange regulatory cofactor NHE-RF2	86.05	10
357	Q16236	Nuclear factor erythroid 2-related factor 2	43.80	12
358	Q16629	Serine/arginine-rich splicing factor 7	58.82	5
359	Q53HL2	Borealin	38.21	4
360	Q6VMQ6	Activating transcription factor 7-interacting protein 1	82.44	27
361	Q7RTN6	STE20-related kinase adapter protein alpha	18.56	4
362	Q86VN1	Vacuolar protein-sorting-associated protein 36	11.92	4
363	Q92585	Mastermind-like protein 1	95.08	23
364	Q96I25	Splicing factor 45	61.00	9
365	Q99081	Transcription factor 12	93.26	20
366	Q9HAU4	E3 ubiquitin-protein ligase SMURF2	19.92	7
367	Q9NQS7	Inner centromere protein	84.53	16
368	Q9UBU9	Nuclear RNA export factor 1	22.65	5
369	Q9UKV5	E3 ubiquitin-protein ligase AMFR	29.55	5
370	Q9UQF2	C-Jun-amino-terminal kinase-interacting protein 1	55.84	10
371	Q9Y263	Phospholipase A-2-activating protein	9.43	4
372	Q9Y3B4	Splicing factor 3B subunit 6	1.60	0
373	Q9Y3E7	Charged multivesicular body protein 3	37.10	4
374	O00268	Transcription initiation factor TFIID subunit 4	76.41	22
375	O00571	ATP-dependent RNA helicase DDX3X	26.02	6
376	P24588	A-kinase anchor protein 5	78.45	9
377	P25490	Transcriptional repressor protein YY1	74.40	8
378	P36956	Sterol regulatory element-binding protein 1	35.48	12
379	P43686	26S protease regulatory subunit 6B	0.24	0
380	P51587	Breast cancer type 2 susceptibility protein	20.66	36
381	P52298	Nuclear cap-binding protein subunit 2	3.87	0
382	P84243	Histone H3.3	33.33	1
383	Q02078	Myocyte-specific enhancer factor 2A	81.85	9

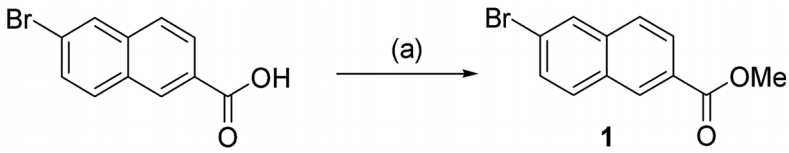
SN	UniProt ID	Protein name	Disorder (%)	BR frequency
384	Q06609	DNA repair protein RAD51 homolog 1	7.69	0
385	Q12772	Sterol regulatory element-binding protein 2	20.16	5
386	Q13177	Serine/threonine-protein kinase PAK 2	38.05	8
387	Q13185	Chromobox protein homolog 3	41.21	5
388	Q15637	Splicing factor 1	90.75	17
389	Q16514	Transcription initiation factor TFIID subunit 12	40.99	3
390	Q86YC2	Partner and localizer of BRCA2	49.66	23
391	Q96KQ7	Histone-lysine N-methyltransferase EHMT2	41.60	12
392	Q9BZI7	Regulator of nonsense transcripts 3B	75.78	9
393	Q9HAU5	Regulator of nonsense transcripts 2	42.77	20
394	Q9HD42	Charged multivesicular body protein 1a	26.53	3
395	Q9Y297	F-box/WD repeat-containing protein 1A	3.80	0
396	Q9Y3Y4	Pygopus homolog 1	79.47	7
397	O00401	Neural Wiskott-Aldrich syndrome protein	74.01	9
398	O43312	Metastasis suppressor protein 1	65.43	16
399	O43516	WAS/WASL-interacting protein family member 1	100.00	11
400	O60673	DNA polymerase zeta catalytic subunit	40.13	56
401	O94916	Nuclear factor of activated T-cells 5	73.42	43
402	O95630	STAM-binding protein	16.51	1
403	P00533	Epidermal growth factor receptor	12.65	6
404	P01133	Pro-epidermal growth factor	9.45	4
405	P02774	Vitamin D-binding protein	0.66	0
406	P06396	Gelsolin	21.72	10
407	P11532	Dystrophin	19.13	23
408	P25791	Rhombotin-2	1.27	0
409	P45983	Mitogen-activated protein kinase 8	18.50	5
410	P49137	MAP kinase-activated protein kinase 2	15.00	3
411	P49450	Histone H3-like centromeric protein A	33.57	1
412	P52565	Rho GDP-dissociation inhibitor 1	19.21	1
413	P61586	Transforming protein RhoA	16.84	4
414	P83916	Chromobox protein homolog 1	57.30	4
415	Q07912	Activated CDC42 kinase 1	46.72	20
416	Q13153	Serine/threonine-protein kinase PAK 1	46.32	10
417	Q14118	Dystroglycan	39.42	6
418	Q14118	Dystroglycan	54.13	6
419	Q14155	Rho guanine nucleotide exchange factor 7	25.16	8
420	Q16539	Mitogen-activated protein kinase 14	3.62	0
421	Q7Z589	Protein EMSY	65.81	28
422	Q86U70	LIM domain-binding protein 1	34.39	3
423	Q8NCD3	Holliday junction recognition protein	55.61	18
424	Q9UBZ9	DNA repair protein REV1	37.81	20
425	Q9UI95	Mitotic spindle assembly checkpoint protein MAD2B	0.00	0
426	Q9UJM3	ERBB receptor feedback inhibitor 1	47.72	10
427	Q9ULV8	E3 ubiquitin-protein ligase CBL-C	19.62	2
428	O15164	Transcription intermediary factor 1-alpha	43.71	9
429	P01135	Protransforming growth factor alpha	2.19	0
430	P01892	HLA class I histocompatibility antigen, A-2 alpha chain	20.82	1
431	P04626	Receptor tyrosine-protein kinase erbB-2	20.03	6
432	P05067	Amyloid beta A4 protein	38.91	13
433	P06748	Nucleophosmin	60.88	7

SN	UniProt ID	Protein name	Disorder (%)	BR frequency
434	P08253	72 kDa type IV collagenase	0.73	0
435	P09429	High mobility group protein B1	68.69	5
436	P14923	Junction plakoglobin	8.72	1
437	P19878	Neutrophil cytosol factor 2	24.33	8
438	P22681	E3 ubiquitin-protein ligase CBL	51.10	16
439	P31947	14-3-3 protein sigma	18.15	3
440	P46100	Transcriptional regulator ATRX	58.07	40
441	P51449	Nuclear receptor ROR-gamma	30.89	4
442	P53999	Activated RNA polymerase II transcriptional coactivator p15	73.02	3
443	P60604	Ubiquitin-conjugating enzyme E2 G2	6.06	0
444	P60953	Cell division control protein 42 homolog	9.57	2
445	P61024	Cyclin-dependent kinases regulatory subunit 1	2.56	0
446	P61968	LIM domain transcription factor LMO4	3.64	0
447	P62136	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	6.99	0
448	P63000	Ras-related C3 botulinum toxin substrate 1	0.00	0
449	P81274	G-protein-signaling modulator 2	20.03	10
450	Q00403	Transcription initiation factor IIB	10.44	1
451	Q05066	Sex-determining region Y protein	32.35	5
452	Q09028	Histone-binding protein RBBP4	23.11	4
453	Q13094	Lymphocyte cytosolic protein 2	67.17	10
454	Q13191	E3 ubiquitin-protein ligase CBL-B	52.04	19
455	Q13227	G protein pathway suppressor 2	91.44	7
456	Q13309	S-phase kinase-associated protein 2	9.91	2
457	Q14839	Chromodomain-helicase-DNA-binding protein 4	40.69	24
458	Q1MX18	Protein inscuteable homolog	2.94	0
459	Q8IX07	Zinc finger protein ZFPM1	67.69	20
460	Q92900	Regulator of nonsense transcripts 1	22.59	9
461	Q96QT6	PHD finger protein 12	47.11	13
462	Q96T88	E3 ubiquitin-protein ligase UHRF1	27.74	11
463	Q9HCE7	E3 ubiquitin-protein ligase SMURF1	23.25	7
464	Q9NRG4	N-lysine methyltransferase SMYD2	0.46	0
465	Q9UBU8	Mortality factor 4-like protein 1	32.87	5
466	Q9UHR4	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1	38.16	10
467	Q9UIS9	Methyl-CpG-binding domain protein 1	45.79	8
468	Q9ULH1	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1	37.73	10
469	Q9UN37	Vacuolar protein sorting-associated protein 4A	18.81	6
470	Q9Y468	Lethal(3)malignant brain tumor-like protein 1	41.76	5
471	Q9Y6K1	DNA (cytosine-5)-methyltransferase 3A	34.32	9

APPENDIX II

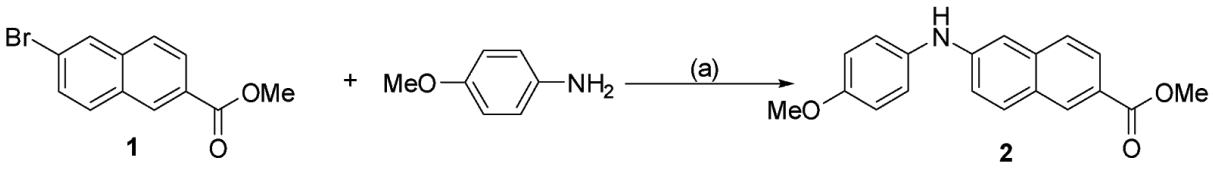
Refers to: Chapter 4

SYNTHESIS OF COMPOUND 5



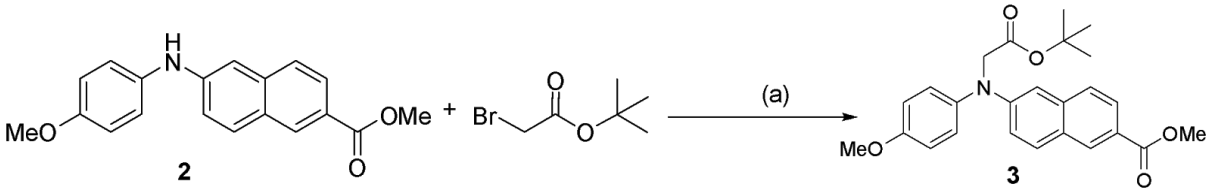
Scheme S4-1: Reagent and conditions: (a) Conc. H_2SO_4, 0 °C to rt, 6 hrs

To a stirred solution of 6-bromo-2-napthoic acid (1 equiv.) in methanol was added H_2SO_4 (10 % by mass) at 0 °C. The reaction was then refluxed for 6 hours. The reaction mixture was then neutralized by 1 M NaOH solution and extracted with ethyl acetate. The ethyl acetate was then concentrated to dryness, and the residue was purified by column chromatography (hexane/ ethyl acetate) to afford **1**.



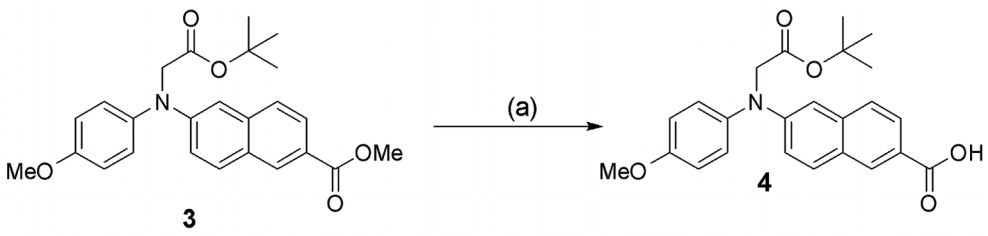
Scheme S4-2: Reagent and conditions: (a) palladium (II) acetate (0.05 equiv.), xantphos (0.1 equiv.) and cesium carbonate (3 equiv.), 80 °C, 4 hrs.

To a stirred solution of compound **1** (1 equiv.) and 4-methoxyaniline (1.2 equiv.) in 1,4-dioxane was added palladium(II) acetate (0.05 equiv.), xantphos (0.1 equiv.) and cesium carbonate (3 equiv.). The reaction was then continued at 80 °C for 4 hours. The reaction mixture was filtered, concentrated to dryness, and the residue was purified by column chromatography (hexane/ ethyl acetate) to afford **2**.



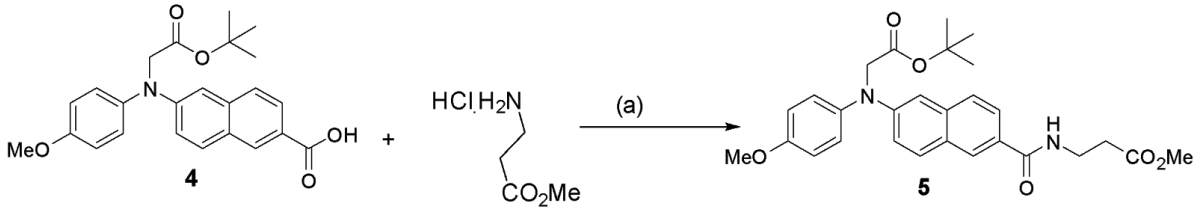
Scheme S4-3: Reagent and conditions: (a) potassium tert-butoxide (1.2 equiv.), DMF, 0 °C to rt, 12 hrs.

To a stirred solution of compound **2** (1 equiv.) in dimethyl formamide was added tert-butyl bromo acetate (1.2 equiv.) and potassium tert-butoxide (1.2 equiv.) at 0 °C. The reaction was then continued at room temperature for 12 hours. The reaction mixture was worked up with ethyl acetate and water. The ethyl acetate was concentrated to dryness, and the residue was purified by column chromatography (hexane/ ethyl acetate) to afford **3**.



Scheme S4-4: Reagent and conditions: (a) Lithium hydroxide (3 equiv.), MeOH-water (5:1), r.t., 2 hrs.

To a stirred solution of compound **3** (1 equiv.) in MeOH-water (5:1) was added lithium hydroxide (3.0 equiv.) at 0 °C. The reaction was then continued at room temperature for 1.5 hours. MeOH was then evaporated and the solution was then neutralized by using 1 M HCl solution and the compound was then extracted with ethyl acetate. The ethyl acetate was then concentrated to dryness to afford **4**, which was carried out to the next step without further purification.



Scheme S4-5: Reagent and conditions: (a) EDC.HCl (1.5 equiv.), HOBT (1.2 equiv.), TEA (3 equiv.), 0 °C to rt, 1.5 hrs.

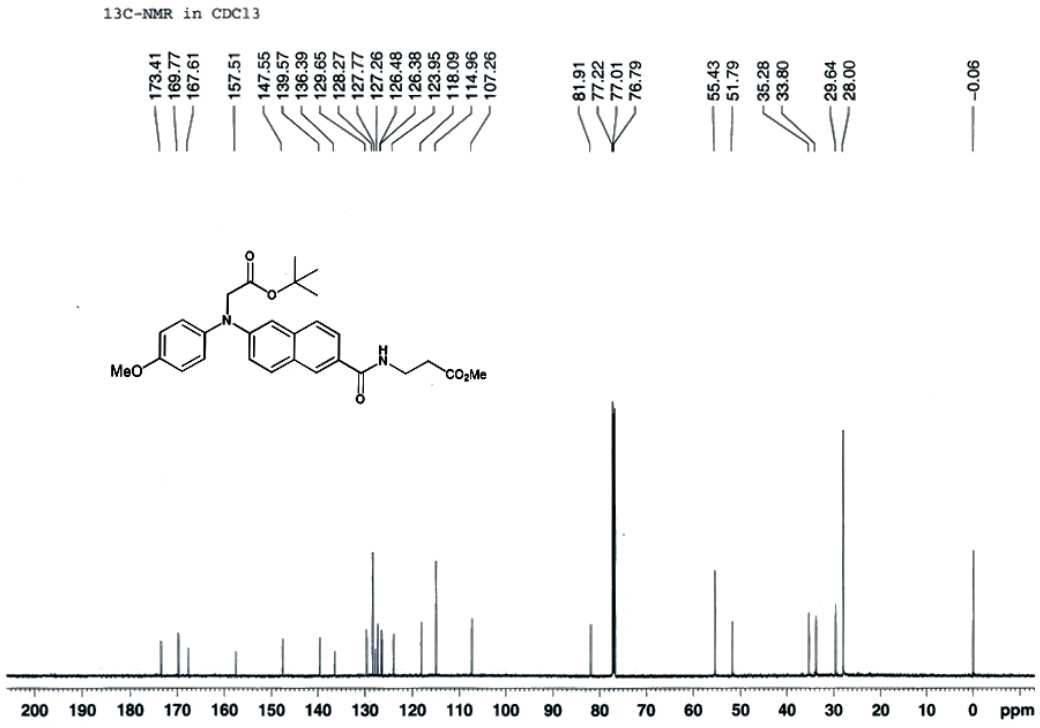
To a stirred solution of compound **4** (1 equiv.) and methyl 3-aminopropanoate hydrochloride (1.2 equiv.) in dry THF was added EDC.HCl (1.5 equiv.), HOBT (1.2 equiv.) and triethyl amine (3 equiv.) at 0 °C. The reaction was then continued at room temperature for 7 hours. The reaction mixture was concentrated to dryness, and the residue was purified by column chromatography (hexane/ ethyl acetate) to afford **5**.

CHARACTERISTIC DATA

Methyl 3-(6-((2-(tert-butoxy)-2-oxoethyl)(4-methoxyphenyl)amino)-2-naphthamido) propanoate: light yellow solid, m.p.= 214 - 215°C, ^{1}H NMR (600 MHz, CDCl3): δ (in ppm) 1.44 (9 H, s), 2.68 (2 H, t, J = 6.0), 3.72 (3 H, s), 3.76-3.75 (2 H, m), 3.83 (3 H, s), 4.37 (2 H, s), 6.85 (1 H, d, J = 2.4), 6.96-6.94 (4 H, m), 7.27-7.26 (2 H, m), 7.57 (1 H, d, J = 8.4), 7.65 (1 H, d, J = 9.0), 7.70 (1 H, d, J= 6.0), 8.12 (1 H, s); ^{13}C NMR (150 MHz, CDCl3): δ (in ppm) 28.00, 31.87, 33.80, 35.28, 51.79, 55.43, 81.91, 107.26, 114.96, 118.09, 123.95, 126.38, 126.48, 127.26, 127.77, 128.27, 129.65, 136.39, 139.57, 144.55, 157.51, 167.61, 169.77, 173.81; HRMS (FAB+): (M+H)/z Calcd for $C_{28}H_{33}N_2O_6$ (M+H)+ 493.2339, Found: (m+H) / z 493.2336.

^{1}H NMR spectra of compound 5 in $CDCl_3$

^{13}C NMR spectra of compound 5 in $CDCl_3$



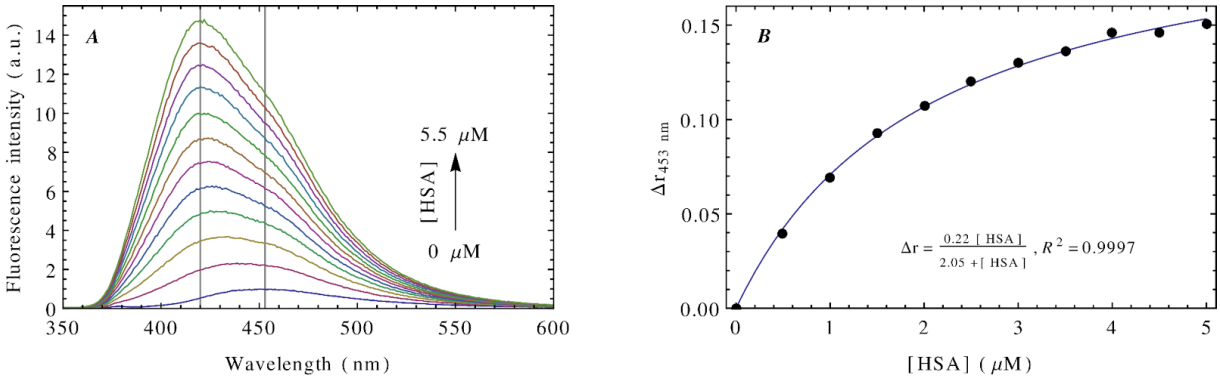


Figure S4-1: Fluorescence emission and anisotropy change of compound **5** in presence of serum albumin. Compound **5** concentration was kept constant at 0.5 μM and the protein concentration was varied from 0 through 5.5 μM. (A) Change in the fluorescence emission spectra of compound **5** as a function of HSA concentration. (B) Change in the compound **5** fluorescence anisotropy with increasing concentration of HSA and the fitted Langmuir isotherm.

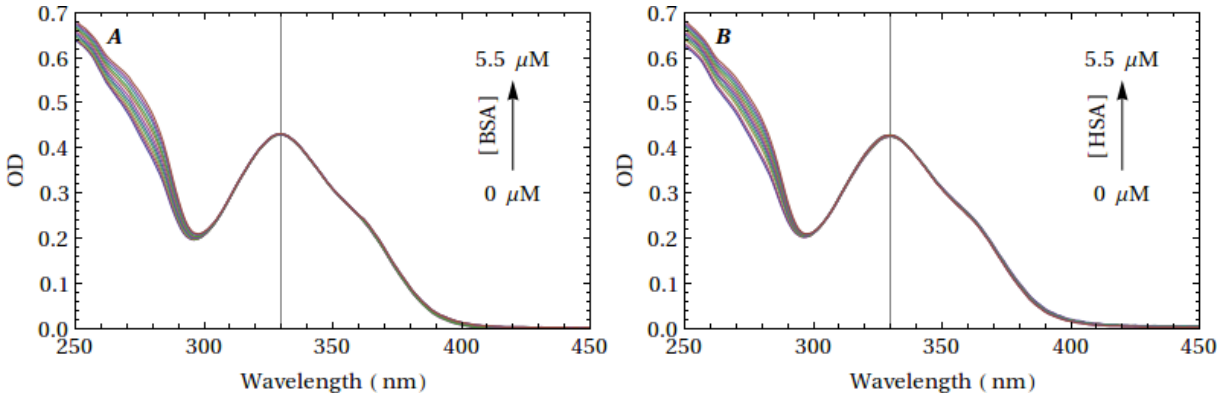


Figure S4-2: Absorption spectra of compound **5** as a function of serum albumins. Compound **5** concentration was kept constant at 0.5 μM and the protein concentration was varied from 0 through 5.5 μM. (A) compound **5** absorption spectra with increasing concentration of BSA. (B) compound **5** absorption spectra with increasing concentration of HSA.

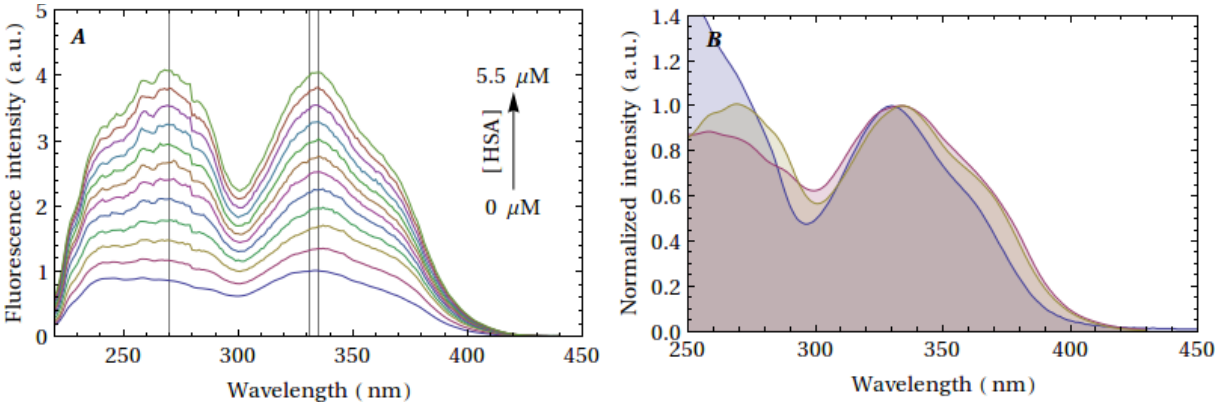


Figure S4-3: Compound **5** binding is an excited state phenomenon. (A) Change in the compound **5** fluorescence excitation spectra upon addition of HSA. (B) Normalized absorption (light blue) and fluorescence excitation spectra of compound **5** (light red) overlapped with compound **5** fluorescence excitation spectra (light yellow) in presence of HSA.

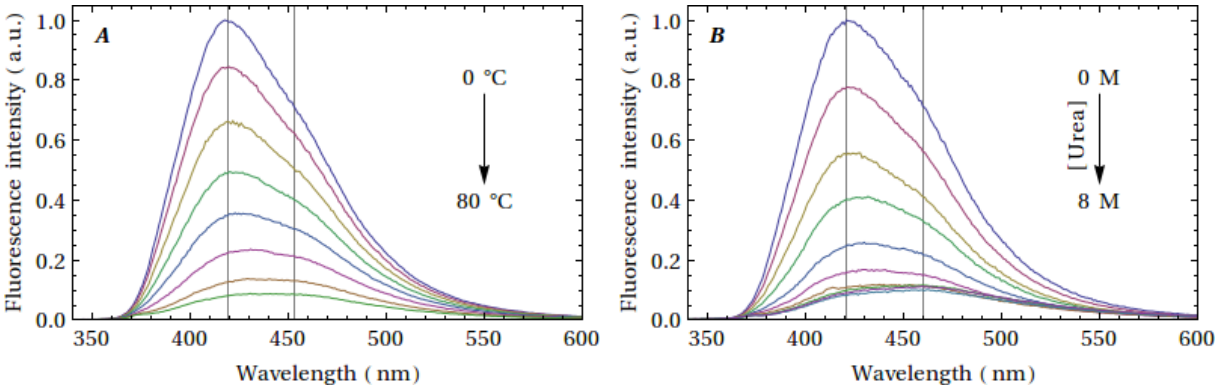


Figure S4-4: Compound **5** interaction with denaturing HSA. (A) Change in the fluorescence of HSA compound **5** complex with the increasing temperature. (B) Change in the fluorescence of HSA compound **5** complex with the increasing concentration of Urea.

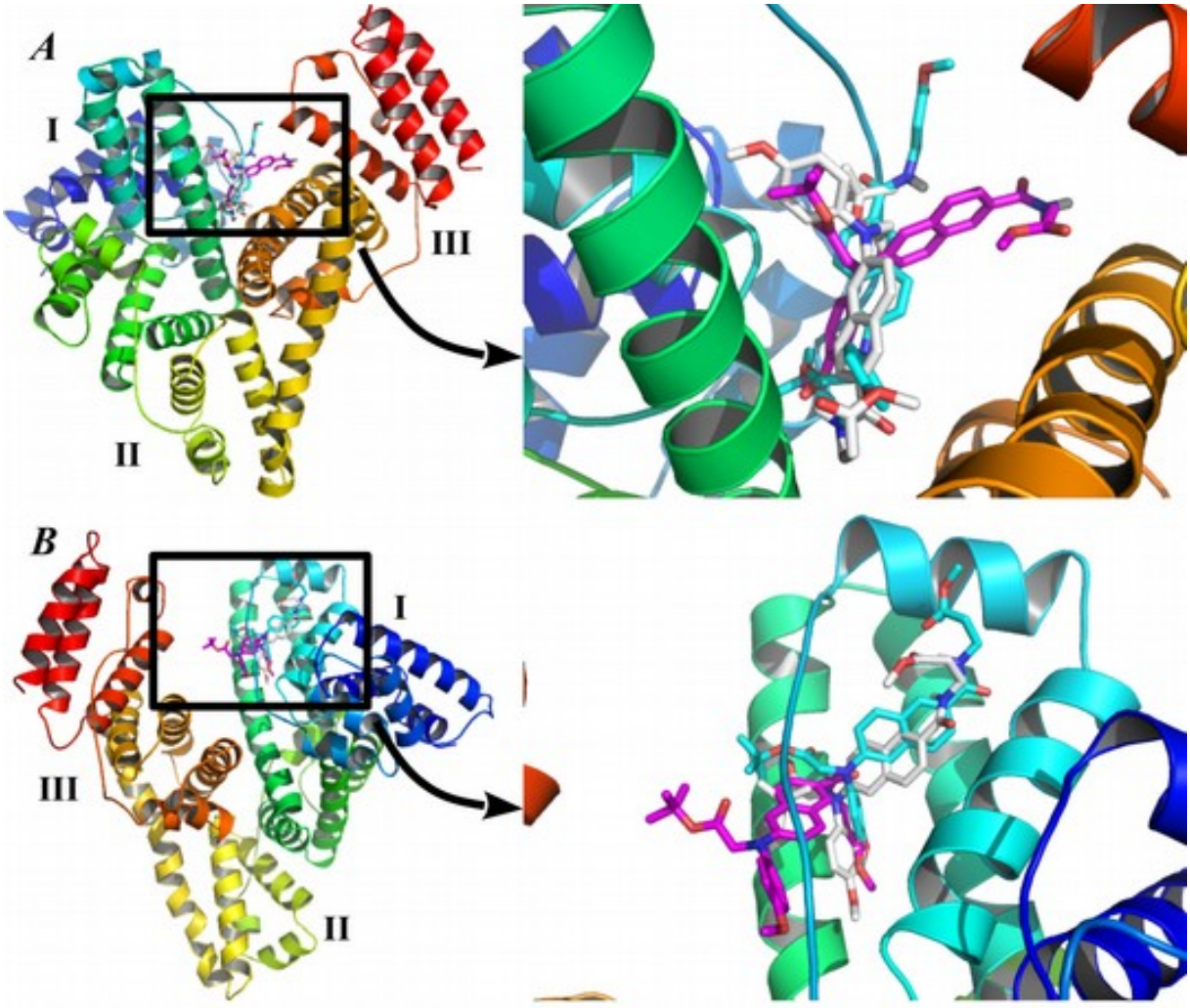


Figure S4-5: Interaction of compound 5 with serum albumins as obtained by three different molecular docking algorithms. AutoDock 4.2, AutoDock Vina and SwissDock results are painted in green, white and cyan, respectively. (A) Best binding conformation of compound **5** with BSA and the close up view. (B) Best binding conformations of compound **5** with HSA; it is also shown in close up. Proteins are shown in ribbon diagram and the ligands in stick model. The three domains of serum albumin are marked with I—III. Standard color representation is used to denote the elements, H, N and O in the ligand.

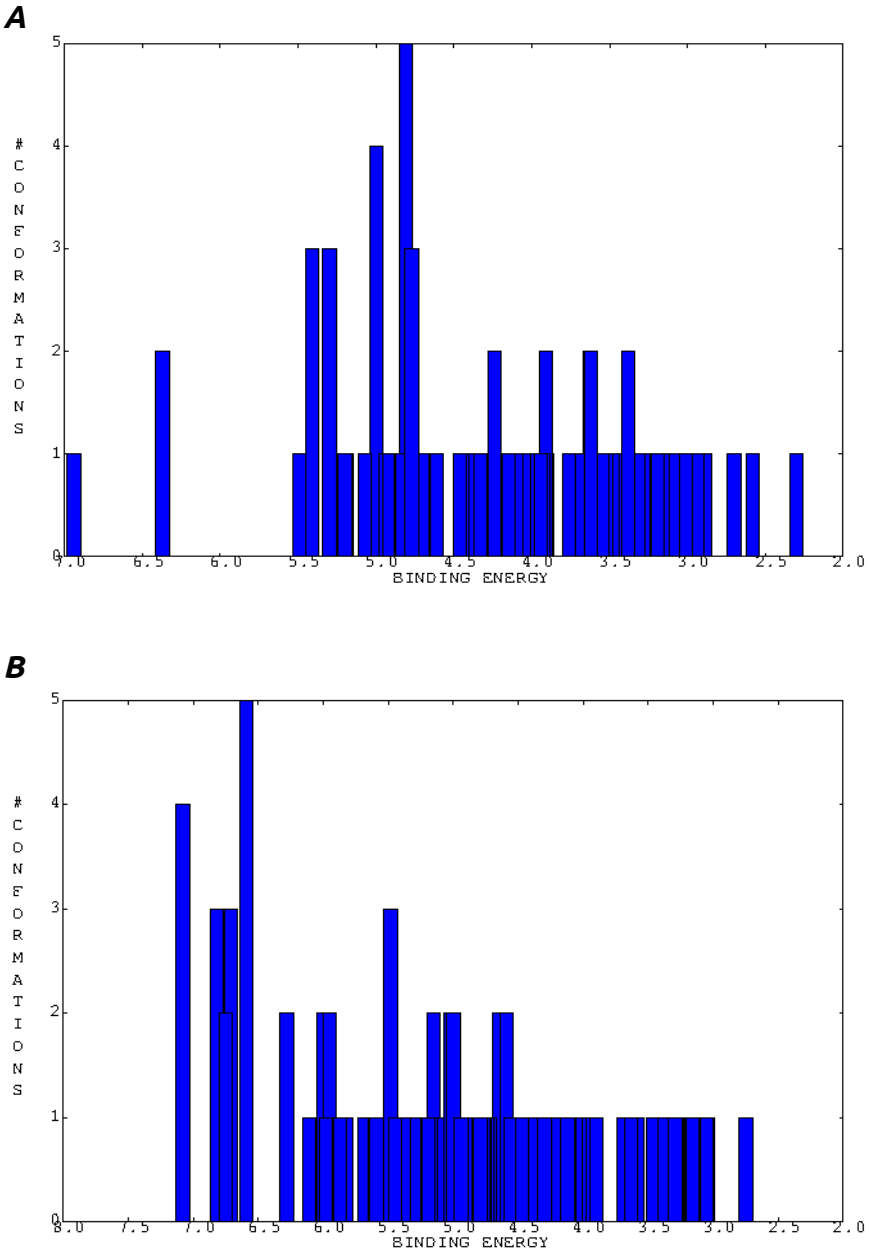


Figure S4-6: Cluster frequency distribution of bound conformations obtained from AutoDock 4.2. All the best conformers within 2 Å standard deviation and 0.5 kcal mol^{-1} energy deviations were grouped together. (A) Cluster of BSA bound conformers. (B) Cluster of HSA bound conformers. Binding energy values are negative.

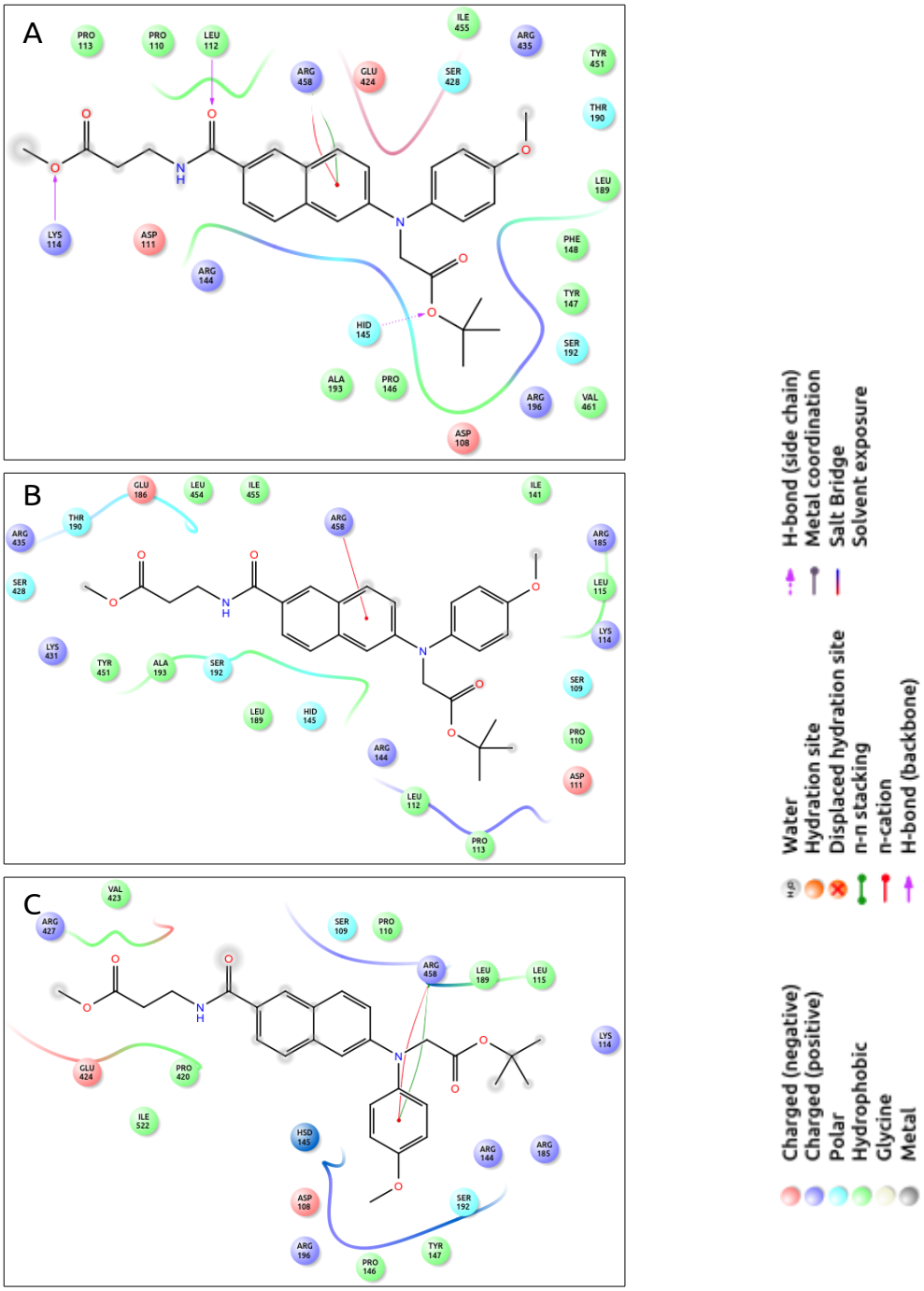


Figure S4-7: Interacting residues of BSA with the compound **5** as obtained by molecular docking experiments. (A) Interacting residues and the types of interaction with BSA as obtained by AutoDock 4.2. (B) Interacting residues and the types of interaction with BSA as obtained by AutoDock Vina. (C) Interacting residues and the types of interaction with BSA as obtained by SwissDock. Color codes and the symbolic expressions for different kind of interactions are mentioned.

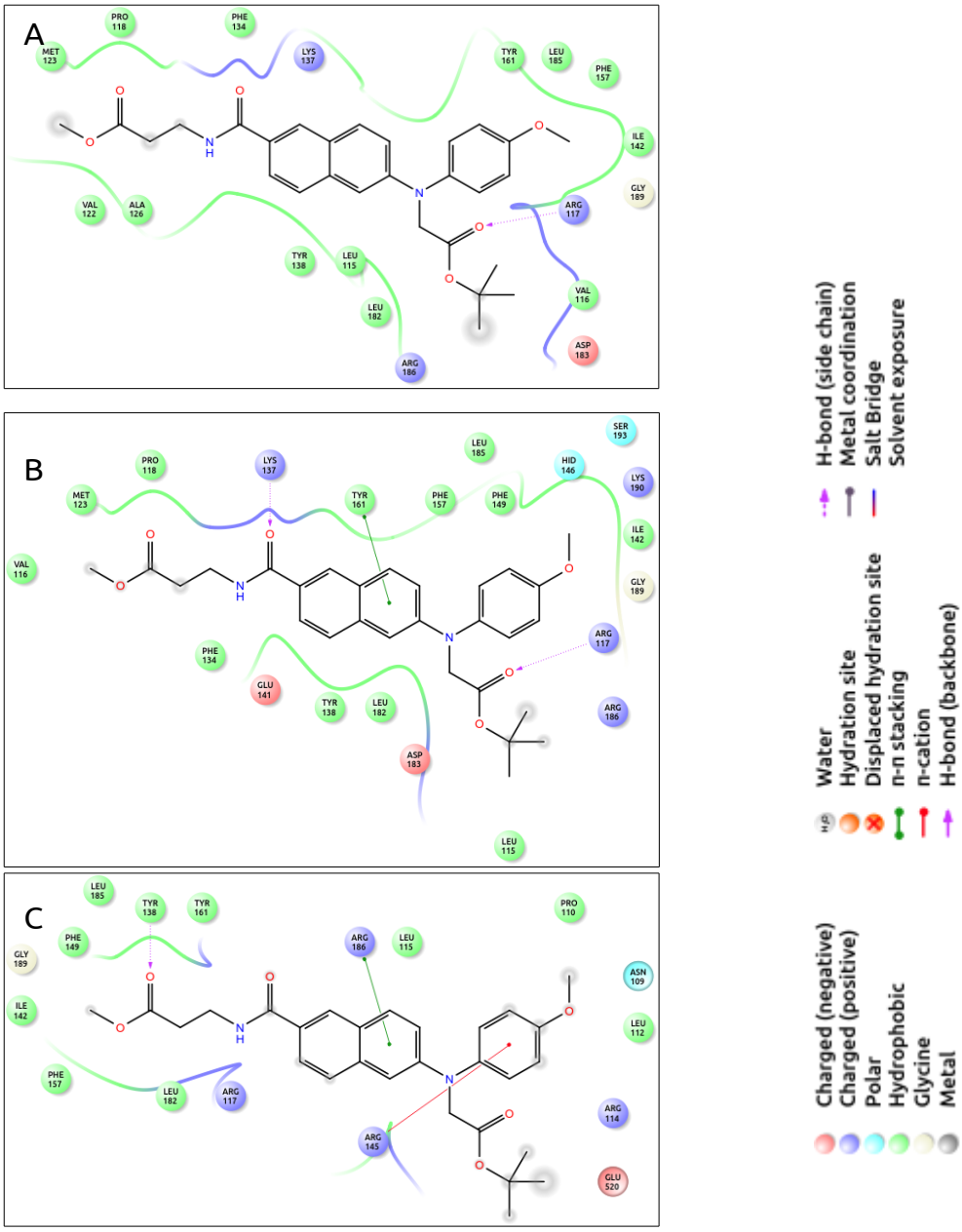


Figure S4-8: Interacting residues of HSA with the compound **5** as obtained by molecular docking experiments. (A) Interacting residues and the types of interaction with HSA as obtained by AutoDock 4.2. (B) Interacting residues and the types of interaction with HSA as obtained by AutoDock Vina. (C) Interacting residues and the types of interaction with HSA as obtained by SwissDock. Color codes and the symbolic expressions for different kind of interactions are mentioned.

APPENDIX III

Refers to: Chapter 5

Optimized coordinates for *anti*-HFIP

E(RB3LYP) = -789.845980266 A.U.

Standard orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	Y	Z
1	6	0	0.134179	-0.081549	1.296144
2	6	0	0.019048	0.749280	0.000000
3	1	0	0.852556	1.457491	0.000000
4	6	0	0.134179	-0.081549	-1.296144
5	8	0	-1.164099	1.496976	0.000000
6	1	0	-1.929007	0.900889	0.000000
7	9	0	1.264237	-0.812276	-1.343656
8	9	0	-0.917038	-0.923996	-1.432919
9	9	0	0.134179	0.744534	-2.360733
10	9	0	1.264237	-0.812276	1.343656
11	9	0	-0.917038	-0.923996	1.432919
12	9	0	0.134179	0.744534	2.360733

Optimized coordinates for *gauche*-HFIP

E(RB3LYP) = -789.844202173 A.U.

Standard orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	Y	Z
1	6	0	1.284454	-0.161619	-0.023570
2	6	0	-0.003859	0.554151	-0.485916
3	1	0	-0.000500	0.511782	-1.583476
4	6	0	-1.301113	-0.138150	-0.030187
5	8	0	-0.054760	1.864175	0.008828
6	1	0	0.653349	2.388406	-0.389436
7	9	0	-1.303896	-1.437608	-0.400668
8	9	0	-1.478557	-0.080596	1.298279
9	9	0	-2.358641	0.457465	-0.619253
10	9	0	1.447519	-1.346313	-0.643942
11	9	0	1.340016	-0.365294	1.300709
12	9	0	2.343374	0.623470	-0.363978

Optimized coordinates for *anti*-HFIP-TMA complex

E(RB3LYP) = -964.356937600 A.U.

Standard orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	Y	Z
1	6	0	-1.417821	-1.259221	-0.010828
2	6	0	-1.060352	0.010699	-0.816860
3	1	0	-1.785816	0.053207	-1.637848
4	6	0	-1.258256	1.319042	-0.018292
5	8	0	0.208784	-0.071030	-1.363077
6	9	0	-2.517986	1.458170	0.447216
7	9	0	-0.418408	1.404749	1.042976
8	9	0	-1.003227	2.378017	-0.815294
9	9	0	-2.683687	-1.234301	0.460115
10	9	0	-0.591872	-1.446326	1.046497
11	9	0	-1.304304	-2.344425	-0.805016
12	1	0	0.962030	-0.072126	-0.684188
13	6	0	3.025960	1.311624	-0.107566
14	1	0	2.410784	1.998073	0.479953
15	1	0	4.073964	1.404407	0.224484
16	1	0	2.964207	1.606397	-1.158749
17	6	0	2.530540	-0.463915	1.461916
18	1	0	2.115892	-1.470274	1.558743
19	1	0	3.551336	-0.460839	1.880682
20	1	0	1.910438	0.221228	2.045013
21	6	0	3.295069	-1.001014	-0.770994
22	1	0	3.234112	-0.707318	-1.822420
23	1	0	4.356681	-1.030158	-0.472509
24	1	0	2.875448	-2.005629	-0.669860
25	7	0	2.520737	-0.059283	0.049801

Optimized coordinates for *gauche*-HFIP-TMA complex

E(RB3LYP) = -964.356549489 A.U.

Standard orientation:

Center Number	Atomic Number	Atomic Type	Coordinates (Angstroms) X	Y	Z
1	6	0	-1.105736	1.351356	0.016424
2	6	0	-0.666925	-0.122753	0.174265
3	1	0	-0.302362	-0.212547	1.207776
4	6	0	-1.831092	-1.125836	0.055354
5	8	0	0.287626	-0.448812	-0.776450
6	9	0	-2.816252	-0.855876	0.944740
7	9	0	-2.379395	-1.147493	-1.171378
8	9	0	-1.377563	-2.370728	0.327829
9	9	0	-1.968642	1.736015	0.984615
10	9	0	-1.673257	1.612933	-1.172552
11	9	0	-0.007117	2.144543	0.129132
12	1	0	1.220350	-0.339942	-0.406324
13	6	0	3.444508	-1.553291	-0.163523
14	1	0	4.533868	-1.559864	0.008674
15	1	0	2.977269	-2.263784	0.523962
16	1	0	3.248776	-1.886365	-1.186207
17	6	0	3.430314	0.751852	-0.913001
18	1	0	3.239663	0.414561	-1.935258
19	1	0	2.947332	1.722790	-0.775666
20	1	0	4.517965	0.873496	-0.777307
21	6	0	3.055283	0.236450	1.422350
22	1	0	4.123005	0.335007	1.681410
23	1	0	2.574552	1.208897	1.560347
24	1	0	2.600074	-0.481062	2.111538
25	7	0	2.865773	-0.217416	0.037994

"Your Highness, I have no need of this hypothesis."
Pierre Laplace (1749-1827), to Napoleon on why his works on celestial mechanics make no mention of God.